

M³CV: A multi-subject, multi-session, and multi-task database for EEG-based biometrics challenge



Gan Huang ^{a,b}, Zhenxing Hu ^{a,b}, Weize Chen ^{a,b}, Shaorong Zhang ^{a,b}, Zhen Liang ^{a,b}, Linling Li ^{a,b}, Li Zhang ^{a,b}, Zhiguo Zhang ^{c,d,e,*}

^a School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, Guangdong, 518060, China

^b Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen, Guangdong, 518060, China

^c Institute of Computing and Intelligence, Harbin Institute of Technology, Shenzhen, China

^d Peng Cheng Laboratory, Shenzhen, Guangdong, 518055, China

^e Marshall Laboratory of Biomedical Engineering, Shenzhen, Guangdong, 518060, China

ABSTRACT

EEG signals exhibit commonality and variability across subjects, sessions, and tasks. But most existing EEG studies focus on mean group effects (commonality) by averaging signals over trials and subjects. The substantial intra- and inter-subject variability of EEG have often been overlooked. The recently significant technological advances in machine learning, especially deep learning, have brought technological innovations to EEG signal application in many aspects, but there are still great challenges in cross-session, cross-task, and cross-subject EEG decoding. In this work, an EEG-based biometric competition based on a large-scale M³CV (A Multi-subject, Multi-session, and Multi-task Database for investigation of EEG Commonality and Variability) database was launched to better characterize and harness the intra- and inter-subject variability and promote the development of machine learning algorithm in this field. In the M³CV database, EEG signals were recorded from 106 subjects, of which 95 subjects repeated two sessions of the experiments on different days. The whole experiment consisted of 6 paradigms, including resting-state, transient-state sensory, steady-state sensory, cognitive oddball, motor execution, and steady-state sensory with selective attention with 14 types of EEG signals, 120000 epochs. Two learning tasks (identification and verification), performance metrics, and baseline methods were introduced in the competition. In general, the proposed M³CV dataset and the EEG-based biometric competition aim to provide the opportunity to develop advanced machine learning algorithms for achieving an in-depth understanding of the commonality and variability of EEG signals across subjects, sessions, and tasks.

1. Introduction

Since the discovery of electroencephalography (EEG) by Hans Berger in 1924 (Berger, 1929), EEG has evolved for use in a wide range of applications (Alotaiby et al., 2014; Cahn and Polich, 2006; Dietrich and Kanso, 2010; Wolpaw et al., 1991) for almost one hundred years. Typically, event-related potentials (ERPs) study focuses on the significant common or mean effects of a cohort, in which the intra- and inter-subject variability are treated as noise and are filtered out by averaging over trials and subjects (Seghier and Price, 2018). Hence ERPs have been widely used for investigating the neurological functions of sensory, motor, and cognitive processes (Kappenman et al., 2021). However, the conventional practice of ERP analysis takes the mean value of the EEG signals across trials and/or subjects to achieve a higher signal-to-noise ratio, but the group-level commonality may not have a reliable effect on the individual level (Boshra et al., 2019; Fröhner et al., 2019; Hu et al., 2022; Infantilino et al., 2018).

The intra- and inter-subject variability pose a great challenge to the individual level explanation and decoding of the EEG signal. Multiple factors may contribute to intra- and inter-subject variability, in-

cluding different brain structures among subjects, non-stationarity of brain activity, neural processing for different tasks, and some unknown factors (Wei et al., 2021). With the availability and affordability of computational power, the recently significant technological advances in machine learning, especially deep learning, have brought technological innovations to EEG signals application in many aspects (Craik et al., 2019). In clinical and psychiatric studies, machine learning technique has pushed EEG-based diagnosis, prognosis, risk stratification, or treatment monitoring toward a more individualized approach (Olbrich and Conradi, 2016). In the application of brain-computer interface (BCI), intention-encoded EEG can be decoded for the direct interaction between the brain and a computer by single-trial decoding (Pfurtscheller and Neuper, 2001; Wolpaw et al., 2000). For system security, EEG-based biometrics become attractive to the research community for personal identification and verification.

However, the lack of generalizability to different subjects and different sessions limited the use of EEG-based machine learning methods in clinical and psychiatric applications (Sui et al., 2020; Fisher et al., 2018). For example, a BCI decoder usually works well in a single personal session, but performs poorly over time or even fails to be applied

* Corresponding author at: Institute of Computing and Intelligence, Harbin Institute of Technology, Shenzhen, China.
E-mail address: zhiguo.zhang@hit.edu.cn (Z. Zhang).

Table 1

Open access EEG databases with characteristics of multi-session, multi-subject, or multi-task.

Database	# Channels	Equipment	# Sessions	# Subjects	# Tasks	Main Usages
Cho et al. (2017)	64	ActiveTwo/Biosemi	1	52	Rest, MI	Motor-BCI
Goldberger et al. (2000)	64	BCI2000	1	109	Rest, MI, ME	Motor-BCI
Brunner et al., 2008	22	Unknown	2	9	MI	Motor-BCI
Zheng et al., 2019	62	EPOC/ Emotiv	4	15	Watch movie clips	Emotion Recognition
Korczowski et al. (2019)	32	g.USBamp/ Gtec	3	50	P300	P300-BCI
Lee et al. (2019)	62	BrainAmp/ BrainProduct	2	54	Rest, MI, P300, SSVEP	BCI inefficiency
Jeong et al. (2020)	60	actiCHamp/ BrainProduct	3	25	11 different upper-extremity movement tasks	Motor-BCI
Kappenman et al., 2021	30	ActiveTwo/ Biosemi	1	40	Rest, N170, MMN, N2pc, N400, P3, LRP and ERN	ERP studies
Langer et al. (2017)	128	Geodesic Hydrocel system	2	126	Passive: Rest, P-SS, NS, Active: CCD, SL, A-SS	Information processing in the developing brain
Gaspar et al. (2011)	128	ActiveTwo/ Biosemi	10	5	ERPs (faces and noises)	ERP reliability
Kumar et al. (2021)	128	GES400/ EGI	2-5	30	OB, FUW, IBA, MMI, PA, SSVEP, PAV	personal identification
Arnaud-Gonzalez et al. (2021)	14	EPOC/ Emotiv	3	21	Rest, AS, VCx and VFx	personal identification
M ³ CV (2022) [this work]	64	BrainAmp/ BrainProduct	2	106 (95 for 2 sessions)	Rest, VEP, AEP, SEP, SSVEP, SSAEP, SSSEP, P300, ME, and SSVEP-SA	Intra- and inter-subject variability and cross-task analysis

*Abbreviations: Rest – Resting-state, MI – Motor Imagery, ME – Motor Execution, LRP – Lateralized Readiness Potential, ERN – Error-related Negativity, P-SS - Passive Surround Suppression, NS - Natural Stimuli, CCD - Contrast Change Detection, SL - Sequence Learning, A-SS – Active Symbol Searching, OB – OddBall, FUW – Familiar and Unfamiliar Words, IBA- Imagining Binary Answers, PA – Passive Audio, PAV – Passive Audio-Visual, AS – Affective Stimulus VCx and VFx – Visual Evoked Potentials Evoked by Check-board and Flash, VEP – Visual Evoked Potential, AEP – Auditory Evoked Potential, SEP – Somatosensory Evoked Potential, SSVEP – Steady-State Visual Evoked Potential, SSAEP – Steady-State Auditory Evoked Potential, SSSEP – Steady-State Somatosensory Evoked Potential.

to another subject (Krusienski et al., 2011; Satti et al., 2010). Although there are meaningful studies concerning cross-session and cross-subject transfer learning for BCI, the theory of common feature space construction in BCI applications is still being studied (Autthasan et al., 2022; Lotte and Guan, 2011; Rodrigues et al., 2019). Similarly, severe overfitting was observed in the application of EEG-based biometrics in the within-session recognition.

In summary, the traditional ERP analysis and statistical test methods help us identify the group-level commonality of EEG, while the recently developed machine learning techniques provided us with a way to further explore the individual-level variability of EEG. However, the lack of generalizability across subjects, sessions, and tasks is still a major challenge for machine learning in neuroimaging. A large-scale multi-subject, multi-session, multi-task EEG database is highly deserved to support this branch of research. Hence, in this study, we established an M³CV (Multi-subject Multi-session Multi-task Commonality and Variability) EEG database to support cross-subject, cross-session, and cross-task EEG studies.

The M³CV database contains 14 types of EEG tasks in 6 experiment paradigms from 106 healthy young adults, in which 95 subjects completed two experimental sessions repeated on different days. To record the EEG data for as many tasks as possible within the limited recording time and to ensure the data quality of each task, we designed the experiments based on suggestions from five experts in neuroscience and psychology (see Acknowledgments).

The experimental paradigms in each session included the following six paradigms with 14 tasks of EEG signals (bolded text).

- Paradigm 1: Resting-state with eye closed and eye open (EC and EO, each lasting for 2 min).
- Paradigm 2: Transient-state sensory with visual, auditory, and somatosensory stimulation (VEP, AEP, and SEP, each having 60 trials)
- Paradigm 3: Steady-state sensory with steady-state visual, auditory, and somatosensory stimulation (SSVEP 1 min, SSAEP 2 min, and SSSEP 2 min)
- Paradigm 4: P300 with oddball experiment (**target P300** with 30 trials, and **nontarget P300** with 570 trials)
- Paradigm 5: Motor execution with the movement of right foot, right hand, and left hand (FT, RH, and LH, each having 80 trials)

- Paradigm 6: SSVEP with selective attention (**SSVEP-SA**, six classes, each class having 12 trials, each trial lasting for 10 s)

When retrieving existing open-access EEG databases (Table 1), we found few databases with multi-task and multi-session EEG signals. The number of subjects in most databases was less than 50. With some specific research objectives, the number of tasks was typically only one or a small number. Multi-session EEG recordings are even harder to collect because it is more difficult to require all subjects to repeat the experiment after a certain period. We also found that existing databases with some attributes of multi-task or multi-session were mainly used for reliability analysis, BCI, and EEG-based personal identification. Reliability analysis (Gaspar et al., 2011) focused on the cross-session reproducibility of EEG signals, which plays a fundamental role in EEG research. BCI studies (Brunner et al., 2008; Goldberger et al., 2000; Jeong et al., 2020; Korczowski et al., 2019; Kumar et al., 2021; Lee et al., 2019) often require the development of cross-session and cross-subject transfer learning algorithms. EEG-based personal identification techniques (Arnaud-Gonzalez et al., 2021; Kumar et al., 2021) use EEG features to identify certain persons among a large number of samples, which should be robust across tasks and sessions. In addition, SEED IV (Zheng et al., 2019) has four sessions of EEG data for EEG-based emotion studies, Langer et al. (2017) presented a dataset combining electrophysiology and eye-tracking intended as a resource for the investigation of information processing in the developing brain, and Kappenman et al (Kappenman et al., 2021) developed ERP Core with six tasks for ERP teaching studies.

Compared with the existing biometrics, such as iris, fingerprint, and face, the security of EEG signal could potentially provide a more secure biometric method as it is confidential, hard to mimic, and almost impossible to steal (Chan et al., 2018; Bidgoly et al. 2022; Gui et al., 2015). The research on EEG-based biometrics has received increased attention in recent years (Wang et al., 2020; Chan et al., 2018; Debie et al., 2021; Jin et al., 2021; Marcel and Millan, 2007). However, there are many challenges such as temporal permanence and robustness to mental state changes in an EEG-based biometric system (Cahn and Polich, 2006; La Rocca et al., 2013; Maiorana, 2021a; Maiorana and Campisi, 2018). At the same time, there is no open EEG-based biometric competition. The lack of a unified test benchmark and platform hinders the develop-

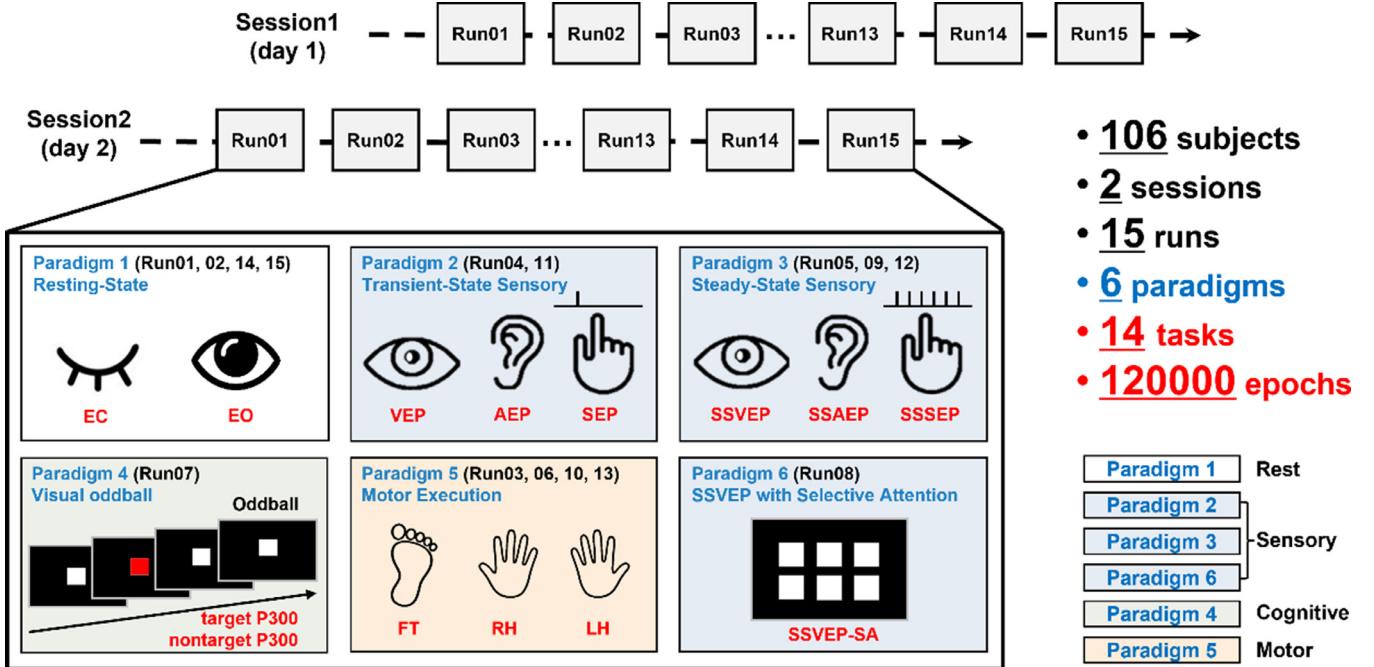


Fig. 1. Illustrations of the experimental paradigm of the M³CV database.

ment of this field. To this end, we decided to open the M³CV dataset to launch an EEG-based biometric competition.

The remainder of the rest of this paper is organized as follows. Session II introduces the M³CV dataset. The details of the EEG-based biometrics competition are presented in Section III. Results are provided in Section IV. The discussion and conclusion are given in Section V.

2. M³CV dataset

2.1. Subjects

A total of 106 healthy subjects from Shenzhen University participated in this experiment. Of these, 95 subjects (Age: 21.3 ± 2.2 years; Gender: 73 males) participated in two sessions of the experiment, which were scheduled on different days, the between-session time is in the range of 6 days to 139 days, with a mean of 20 days. All subjects had normal hearing, normal or corrected-to-normal vision, and no history of neurological injury or disease (as indicated by a self-report). During the experiment, the subjects were seated in comfortable chairs and kept about one meter from the screen.

Ethical approval of the study was obtained from the Medical Ethics Committee, Health Science Center, Shenzhen University (No. 2019053). All subjects were informed of the experimental procedure, and they signed informed consent documents before the experiment, in which they agreed to make their data open to access for research aim on the premise of concealing their personal information.

2.2. Experimental paradigm

Fig. 1 shows that the whole experiment was arranged in two sessions on separate days. The entire experiment with 6 paradigms, 15 runs, and 14 tasks was completed within 2 h (around 50 min of recording time and 70 min for experiment preparation and rest between the consecutive runs).

The description of the 6 experimental paradigms is given below:

- (1) **Resting-state:** Runs 01 and 14 were resting-state EEG signals with eyes-closed (EC), while Runs 02 and 15 were resting-state EEG signals with eyes-open (EO). Each lasting for 1 min. Instead

of the fixation cross on the screen, the subject was asked to keep their eyes fixed on the LED (staying off) during the resting state with their eyes open. In the eyes-open run, the subjects were required to keep their eyes fixed in front and blink as little as possible.

- (2) **Transient-state sensory:** EEG elicited by visual, auditory, and somatosensory stimuli were recorded in Runs 04 and 11. For each run, 30 trials of VEP, AEP, and SEP were arranged in random order. Each stimulus lasted 50 milliseconds, and the inter-stimulus interval (ISI) was set at 2–4 s. On average, each run lasted 4.5 min.
- (3) **Steady-state sensory:** A train of visual, auditory, and somatosensory stimuli were released in Runs 05, 09, and 12 respectively. Considering the different frequency responses of different modality stimuli, and the higher signal-to-noise ratio of SSVEP compared to SSAEP and SSSEP, the stimulation frequencies and recording times were different. They were 10 Hz (Herrmann, 2001) and 1 min for SSVEP in Run 05, 45.38 Hz (Galambos et al., 1981) and 2 min for SSAEP in Run 09, and 22.04 Hz (Snyder, 1992) and 2 min for SSAEP in Run 12.
- (4) **Visual Oddball:** A visual oddball experiment was arranged in Run 07 with the red square as the target stimuli and a white square as the nontarget stimuli on the screen. Each square lasted 80 ms with the ISI 200 ms. Six hundred trials of stimuli were arranged in 2 min, in which target stimuli had a possibility of 5% (30 trials of target stimulation and 570 trials of nontarget stimulation). The subjects were asked to count the number of red squares and report after the run so that their attention would remain on the screen.
- (5) **Motor Execution:** Executed movement was performed in Runs 03, 06, 10, and 13. During these runs, the subjects were instructed to respond to a visual cue by gripping their left hand (LH) or right hand (RH), or by lifting their right ankle (FT) for a duration of 3 s, i.e., until the cue offset. No feedback was provided during the online recording. To ensure their motor areas were being fully activated, the subjects were required to perform the real executed movements of FT, RH, and LH at a rate of twice per second or faster, at approximately 80% of their maximum vol-

Table 2Description of 14 types of EEG tasks recorded in the M³CV database.

No.	Name	Paradigm	Run	Duration	Trigger	Description
1	EC	1	01, 14	2 min	S 1	Resting state with eyes closed
2	EO	1	02, 15	2 min	S 2	Resting state with eyes open
3	VEP	2	04, 11	60 trials	S 3	Visual evoked potential
4	AEP	2	04, 11	60 trials	S 4	Auditory evoked potential
5	SEP	2	04, 11	60 trials	S 5	Somatosensory evoked potential
6	SSVEP	3	05	1 min	S 23	Steady-state visual evoked potential
7	SSAEP	3	09	2 min	S 24	Steady-state auditory evoked potential
8	SSSEP	3	12	2 min	S 25	Steady-state somatosensory evoked potential
9	Target P300	4	07	30 trials	S 9	Epochs with target stimuli in P300 experiment
10	Nontarget P300	4	07	570 trials	S 10	Epochs with nontarget stimuli in P300 experiment
11	FT	5	03, 06, 10, 13	80 trials	S 6	Right ankle movement
12	RH	5	03, 06, 10, 13	80 trials	S 7	Right hand movement
13	LH	5	03, 06, 10, 13	80 trials	S 8	Left hand movement
14	SSVEP-SA	6	08	12 segments/2.5 min	S 11-S 16 (cue) S 17-S 22 (begin)	Steady-state visual evoked potential with selective attention

unitary contraction, while keeping their upper body still. There was no external tool, like a metronome or hint on the screen, to remind the subject, because it may have produced unnecessary evoked potential as an external stimulus. During the experiment, the experimenter monitored the movements of the subjects. If the subjects did not move fast enough or were found that their body moved simultaneously, the experimenter would abandon the recording of the current run, correct the subject's movement and let them practice several more trials. The experimenters continuously monitored whether the movements of the subjects met these standards and corrected them when necessary. To minimize the intra- and inter-subject variability, we tried to ensure the consistency of the movements for all subjects. Hence, the executed movement was used instead of motor imagery in classical BCI experiments, and no feedback or training was given to the subjects.

(6) **SSVEP with selective attention:** Six white squares flashed simultaneously at different frequencies in Run 08. First, the target square turned red for 200 ms. After 500 ms, all squares with the frequency of 7 Hz, 8 Hz, 9 Hz, 11 Hz, 13 Hz, and 15 Hz began to flash and lasted 10 s. The subject was asked to focus on the target square during the 10 s by covert visual attention with a fixation on the middle of the screen. In total, there were 12 trials arranged in Run 08. For each session of the experiment, they were instructed to perform 6 types of experimental paradigms, which were arranged in 15 runs, in which 14 tasks with rest, sensory, cognitive, and motor-related EEG signals were listed in Table 2.

2.3. Experimental platform

The continuous EEG signals were recorded using an EEG amplifier (BrainAmp, Brain Products GmbH, Germany) and multichannel EEG caps (64 Channel, Easycap). The signals were recorded at a sampling rate of 1000 Hz by 64 electrodes, placed in the standard 10–20 positions. The electrodes FCz and AFz served as reference and ground, respectively. Before data acquisition, the contact impedance between the EEG electrodes and the cortex was calibrated to be lower than 20 kΩ to ensure the quality of EEG signals during the experiments.

An Arduino Uno platform was programmed to release the visual, auditory, and somatosensory stimuli in tasks 2 and 3, which communicated with the Matlab program (The MathWorks Inc., Natick, USA) on a PC through a serial port.

- **Visual stimuli** for VEP and SSVEP were delivered by a 3 W light-emitting diode (LED) with a 2 cm diameter circular light shield, which was placed in the center of the visual field of view 45 cm away from the subjects' eyes. The mean LED intensity was 1074 Lux as measured by a light meter (TES-1332A, TES).
- **Auditory stimuli** for AEP and SSAEP were presented via a Nokia WH-102 headphone. 1000 Hz pure tone was applied for the stimuli.

The intensity was set at a comfortable level (75 dB SPL on average) for all subjects as measured by a digital sound level meter (Victor 824, Double King Industrial Holdings Co., Ltd. Shenzhen, China).

- **Somatosensory stimuli** for SEP and SSSEP were generated by a 1027 disk vibration motor. Since there was no effective tool to measure the output intensity of the vibrator directly, we have to report the detail of the product parameters with the rated power of 3 W, the efficiency of 80%, and dimensions of 10mm*2.7mm).

For the other tasks, A 24.5-inch screen (1920*1080) with a 240-Hz refreshing rate (Alienware AW2518H, Miami, USA) was used to present the visual stimuli or cues by programmed using Psychtoolbox-3 (<http://psychtoolbox.org/>) in Matlab. The red and white squares were delivered in a sequence at the center of the screen with a black background for visual oddball. Three white squares were also used in the motor execution paradigm to indicate the movement. Six white squares flashed simultaneously at different frequencies to deliver the stimulus in SSVEP with selective attention paradigm. The size of the square was set to be 300*300 pixels in these paradigms.

2.4. EEG signal pre-processing

The data pre-processing pipeline on the M³CV database is illustrated in Table 3. The raw EEG signals were recorded with BrainVision core data format (each recording consisting of a .vhdr, .vmrk, .eeg file triplet), which were managed with Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016; Pernet et al., 2019). For each recording, the bad channels were interpolated first. Channel FCz (the reference) was added back, and channel IO was removed. Then all signals were filtered by a 0.01–200-Hz band-pass filter and a 50-Hz notch filter. A 2-order Butterworth zero-phase filter was applied in the above two steps of filtering. After re-referencing to TP9/TP10, artifacts produced by eye blinks or eye movements were identified and removed manually by Independent Component Analysis (ICA) (Huang et al., 2020).

During the signal preprocessing, we interpolate bad channels for 22 subjects and one subject was removed due to strong 10 Hz artifacts among the 95 subjects who complete the two sessions. For Motor execution, preventing the interference of EMG artifacts caused by the actual movement was crucial during our data collection, especially in the movement paradigm. Firstly, we asked the subject to keep their torso still in both foot and hand movements. That is also the reason why we set right foot movement instead of two-foot movement in the experiment design. Secondly, the experimenters continuously monitored the subjects' movement and online EEG display to ensure the quality of data collection. Thirdly, since the amplitude of EMG single is much higher than the EEG signal, the contamination of EMG artifacts is highly visible. Also, we did not see strong EMG artifacts during the signal preprocessing. Furthermore, no strong EMG artifact is observed in the time-frequency analysis. The main EEG responses came from C3, C4, and Cz

Table 3
Pipeline for EEG pre-processing.

Software	Matlab 2018b & Letswave7 (Letswave.cn)
Band-pass filtering	Butterworth filter, 0.01-200 Hz, 4 th order, 24 dB/octave, zero-phase
Notch filtering	Butterworth filter, 49-51 Hz, 4 th order, 24 dB/octave, zero-phase
Channel interpolation	Bad channels were identified manually and interpolated with the mean value of the three surrounding channels
Re-reference	Re-reference to the mean value of TP9 and TP10
Artifacts removal by ICA	Eye movement related ICA components were identified by visual inspection of their scalp topographies, time courses, and spectra.

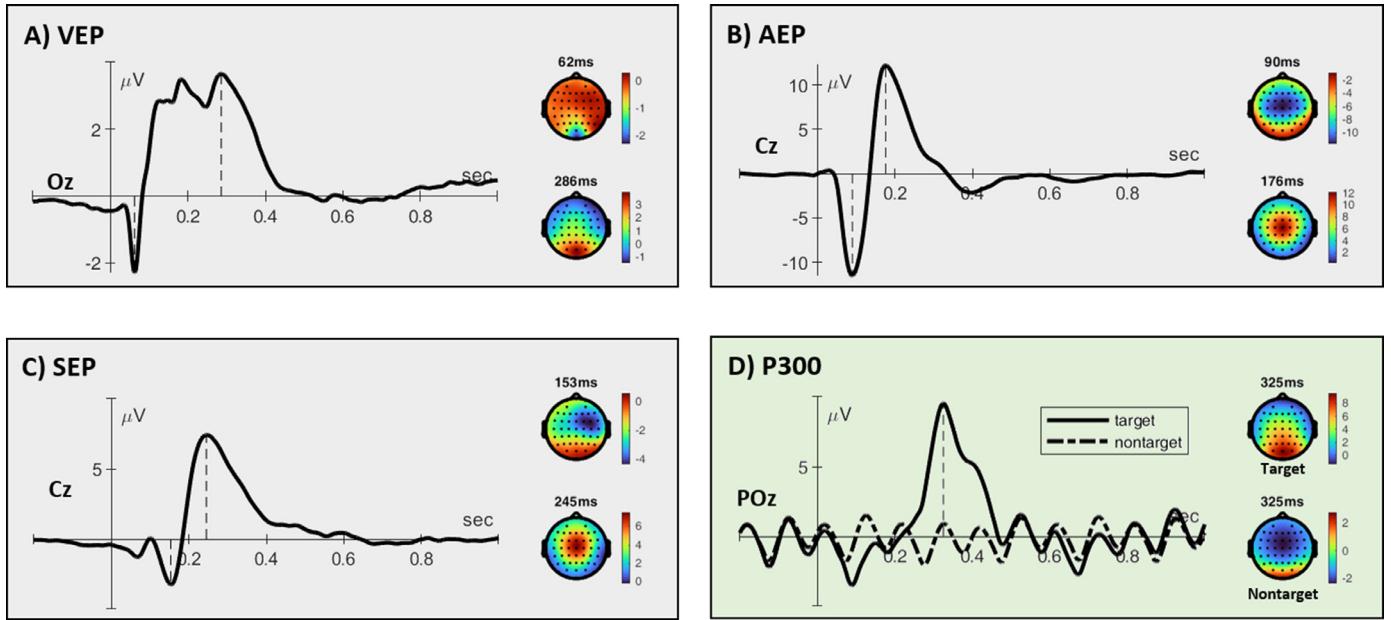


Fig. 2. Time domain EEG results for (A) VEP, (B) AEP 2, (C) SEP in paradigm 2 and (D) P300 in paradigm 4.

at μ and β band. The γ band response was much smaller, which was the main frequency band for EMG artifacts.

It should be mentioned that we did not perform bad epoch rejection in pre-processing. The machine learning task is different from conventional ERP analysis. The robustness to the outlier is also an important feature of the machine learning algorithm. Besides, the criteria for rejecting epochs were different across research groups, thus affecting the repeatability of the algorithm.

2.5. Basic EEG signal visualization

To show the basic characteristic of EEG signals in M³CV, we perform time domain analysis on ERP signals, Frequency domain analysis on resting-state EEG and steady-state evoked potentials, and time-frequency domain analysis on motor-related signals.

2.5.1. Time domain analysis

The ERPs of VEP, SEP, AEP, and target and non-target P300 were analyzed in the time domain. A Butterworth bandpass filter with 0.5–30 Hz was applied to the preprocessed EEG signal. It should be noted that the 0.5–30Hz bandpass filter was used only for ERP data visualization, and the data set provided in the biometrics competition was filtered with a 0.01–200Hz bandpass filter. After segmentation and averaging, baseline correction was performed from −0.5 to 0 s to obtain the ERP for each subject.

For time-domain analysis. The grand-averaged waveforms of VEP, AEP, SEP, and target- and nontarget-P300 are shown in Fig. 2 (A–D), with the topographies at their corresponding peaks. The response of VEP

mainly concentrated on the occipital area. AEP and SEP shared a similar N2/P2 waveform in the central area. A classical N75-P100-N135 complex is not seen here, which is typically observed from pattern-reversal VEP, for which high-contrast, black-and-white checkerboards are used as stimuli. As the vibrator was placed on the subject's left hand, the N2 response of SEP showed a response from the contralateral primary sensory area. The topographies for the P2 components from AEP and SEP are quite similar. In the oddball experiment, the target P300 response was mainly concentrated on the channel POz at around 300 ms. By comparison, the nontarget response mainly came from the occipital area because the visual stimuli were used in the oddball experiment.

2.5.2. Frequency domain analysis

For the task of resting-state with eye closed and eye open, the EEG signals were firstly re-referenced to a common average reference after preprocessing. Welch's method with a 2-s window and 50% overlap was applied. After transforming the total power level into decibels and averaging, the frequency domain responses for resting-state EEG with eye closed and eye open were obtained for each subject. The processing pipeline for EEG signals from steady-state sensory and SSVEP with selective attention tasks was similar. But no re-reference was performed, and TP9/TP10 was still used as the reference. FFT was applied instead of Welch's method to obtain a sharp frequency response for SSVEP, SSAEP, SSSEP, and six classes of SSVEP-SA signals.

For frequency domain analysis of the resting-state EEG with EC and EO (Fig. 3A), the difference was concentrated on the occipital area. With the increase in frequency, the amplitude of the EEG response decreased at a rate of approximately 1/freq for both EC and EO. The main differ-

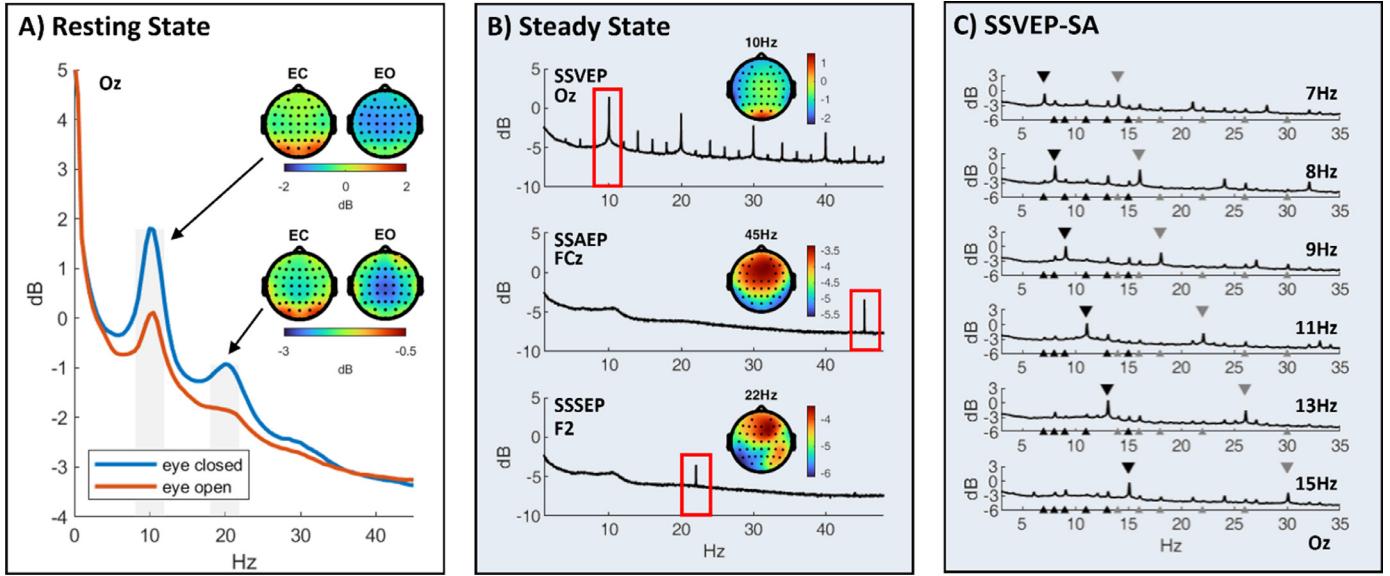


Fig. 3. Frequency domain EEG results for (A) Resting-state EEG with eye closed and eye open, (B) Steady-state Evoked potential with SSVEP, SSAEP, and SSSEP, and (C) SSVEP with selective attention.

ence between EC and EO was the three peaks that appeared in sequence around 10 Hz, 20 Hz, and 30 Hz. For the first peak in the interval of 8–12 Hz, the main response was in the occipital area, the frequency power at Cz (0.45 dB) is larger than C3 (0.17 dB) and C4 (0.21 dB) in the condition of EO, while the frequency power at Cz (-1.05 dB) would be smaller than C3 (-0.88 dB) and C4 (-0.73 dB) in the condition of EC. The topographies for the second peaks in the interval of 18–22 Hz had a similar shape, while the difference mainly came from the magnitude. For SSEP analysis in Fig. 3B, the brain areas and frequency band of the induced SSEP were different for visual, auditory, and somatosensory stimuli, which were around 10 Hz at the occipital area for SSVEP, around 45 Hz at the frontal-central area for SSAEP, and around 22 Hz at the primary sensory area for SSSEP. Fig. 3C shows how attention modulated the foundation frequency (black triangles) and their harmonic (gray triangles) responses at different frequency points. With selective attention, the target frequency (downward-pointing triangles) responses were larger than those at other nontarget frequencies (upward-pointing triangles).

2.5.3. Time-frequency domain analysis

For the motor execution tasks, since the EEG responses were complex in different time intervals and frequency bands, a continuous wavelet transform was applied for the time-frequency domain transform, in which the complex Morlet wavelet was used with the central frequency 1.5 Hz and bandwidth 1 Hz. Considering the computation complexity, the EEG signal was downsampled to 200 Hz to reduce the time for computation. After the continuous wavelet transform, the time-frequency response was further downsampled to 50 Hz to reduce the storage space on the hard disk and computer memory.

Time-frequency analysis was performed for the motor task because foot and hand movements led to rich neural oscillation changes in different time-frequency-spatial regions. Fig. 4 gave a comprehensive illustration of these changes caused by these three types of movements, which are FT, RH, and LH. Further, no strong EMG artifact is observed in the time-frequency analysis. The main EEG responses came from C3, C4, and Cz at μ and β band. The γ band response was much smaller, which was the main frequency band for EMG artifacts.

The corresponding topographies from six regions of interest (ROI) are detailed below.

- ROI #1 (from 0.1 to 0.3 s, 2 to 7 Hz): The response in this region was due to Motor-related Cortical Potential (MRCP), which is a super low-frequency negative shift in EEG recording. MRCP may come earlier than the movement because the cortical processes are employed in the planning and preparation of movement. Hence, MRCP can be used for predicting the movement in BCI applications. In this experiment, subjects did not know which movement and when they were going to execute. Hence the MRCP response came out after the zero point. The brain area of its response was the same as the motor area of the corresponding action.
- ROI #2 (from 1 to 3 s, 9 to 13 Hz): The ERD response from the μ rhythm oscillation was observed in this region. By carefully comparing the responses of RH and LH, we found that the ERD phenomenon happened in motor areas of both left and right hemispheres. But the ERD in the contralateral brain area was larger than that in the ipsilateral brain area.
- ROI #3 (from 1 to 3 s, 20 to 30 Hz): The ERD response from the β rhythm oscillation was also well reported for motor-based BCI application. Similar to ROI #2, the ERD phenomenon happened in the motor areas of both the left and right hemispheres for RH and LH. The interhemispheric difference was mainly located in the parietal-central area in ROI #2 and located in the frontal-central area in ROI #3.
- ROI #4 (from 1 to 3 s, 55 to 90 Hz): During the movement, high γ oscillation also showed an interhemispheric difference for RH and LH. But the magnitude was weaker than that of ROI #2 and ROI #3. We found that the magnitude change in ROI #4 was ERD for both RH and LH, which was ERS for FT. But the magnitude was quite small for the high-frequency EEG recording.
- ROI #5 (from 4 to 4.5 s, 12 to 14 Hz): After the movement, μ rhythm ERS phenomenon occurred in the corresponding motor area for FT, RH, and RL. With the large sample size in our dataset, some more detailed results can be observed. For example, the μ rhythm ERS for FT occurred earlier at a higher frequency band (around 12.5 Hz, and 4.2 s) than ERS for RH and LH (around 9 Hz, and 5 s).
- ROI #6 (from 4 to 4.5 s, 20 to 30 Hz): β rhythm ERS phenomenon was seen in this region. We found that from 3.7 to 4.2 s, the μ rhythm ERD and β rhythm ERS occurred simultaneously for RH and RL. The μ rhythm ERS and β rhythm ERS did not occur at the same time for RH and LH, but occurred at almost the same time interval for FT.

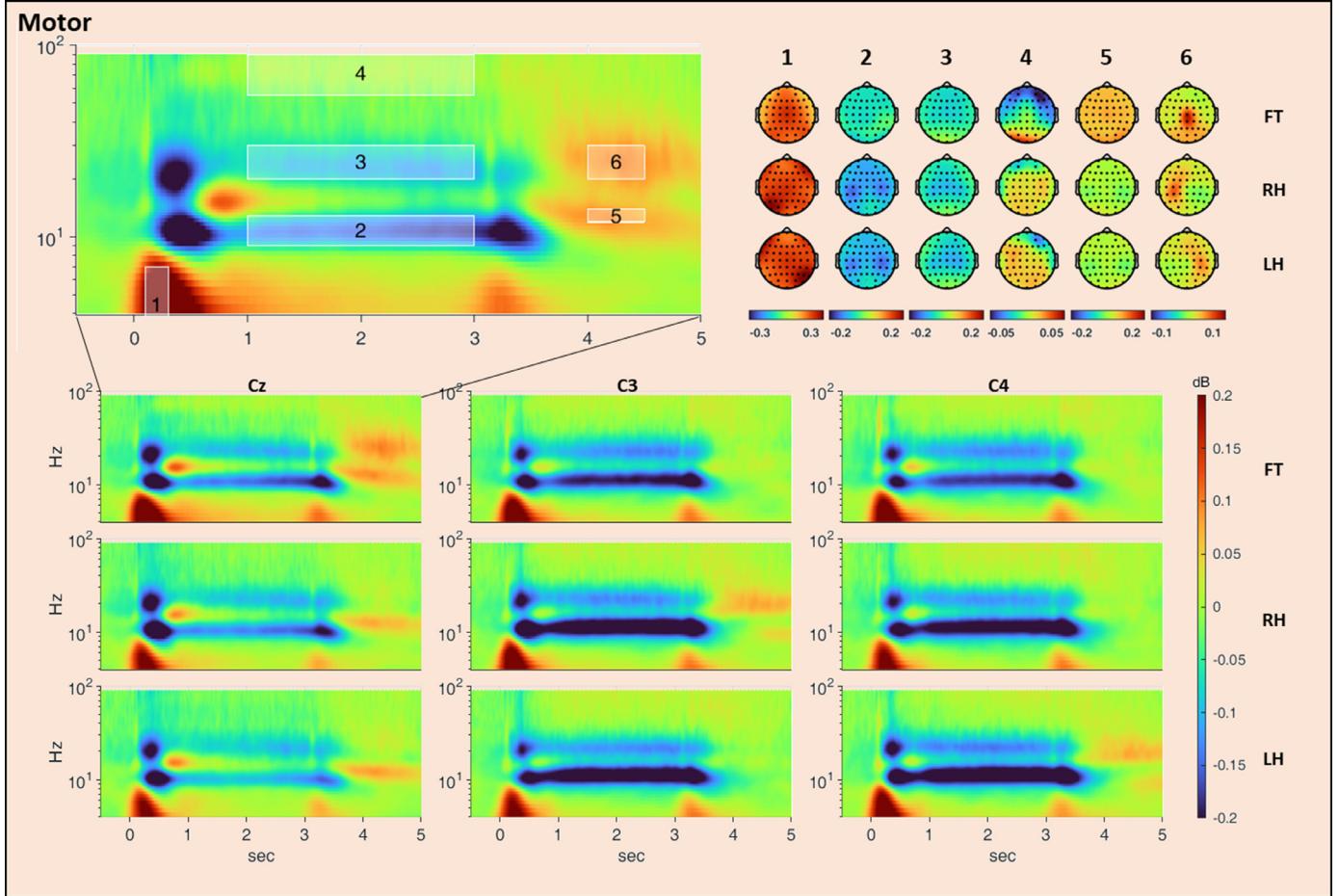


Fig. 4. Time-frequency domain EEG results for motor execution task for FT (right angle movement), RH (angle hand), and LH (left hand movement) at channels Cz, C3, and C4. The topographies of the mean magnitude from the six typical regions are illustrated for three types of movements.

3. EEG-based biometrics

3.1. Literature review

Several studies have reported that EEG-based identification and verification systems are capable of high recognition accuracy (Kong et al., 2019; Koike-Akino et al., 2016). However, many of them have not been evaluated under more rigorous conditions, such as testing in a database with large population sizes, across sessions, and across tasks, which were detailed as follows:

- **Uniqueness:** Although many studies have revealed the uniqueness of EEG signals and their ability to discriminate between subjects, the tested databases normally have a small population size of fewer than 50 subjects (Delpozo-Banos et al., 2018; Arnaud-Gonzalez et al., 2021; Chen et al., 2016). For employing EEG as a biometric trait, it must be tested on a larger database.
- **Permanence:** EEG signals could be different in days even for the same task, which poses a major challenge to decoding subject identity (La Rocca et al., 2013; Maiorana et al., 2016; Maiorana and Campisi, 2018). Further, in single-session settings, machine learning algorithms may recognize the special electrode situation and preprocessing parameters rather than the EEG-based personal trait.
- **Robustness:** For the same session, EEG signals could vary significantly under different tasks or experimental conditions. A good biometric system should be robust to the above conditions (Wang et al., 2019; Maiorana, 2021a; Kumar et al., 2021; Kong et al., 2018), otherwise,

an enrolled subject may be rejected or misidentified by the system when they are under different mental states.

According to the three points mentioned above, we did a short review of the existing studies on multi-session and multi-task EEG-based biometrics as shown in Table 4. Compared with our proposed database, the population size of these studies is quite small, especially for La Rocca et al. (2013) and Zeynali and Seyedarabi (2019). As for feature extraction, hand-crafted features, such as AR, MFCC, and PSD, are still the dominant choice while only Maiorana (2020) and Kumar et al. (2021) applied deep learning methods. In terms of performance metrics, the recognition accuracy and the equal error rate were the most used metric for identification tasks and verification tasks, respectively (Arnaud-Gonzalez et al., 2021; Kumar et al., 2021). It should be noted that these studies were performed on multi-session and multi-task EEG databases, but only Kumar et al. (2021) applied cross-session and cross-task evaluation to their proposed algorithm.

3.2. Learning task

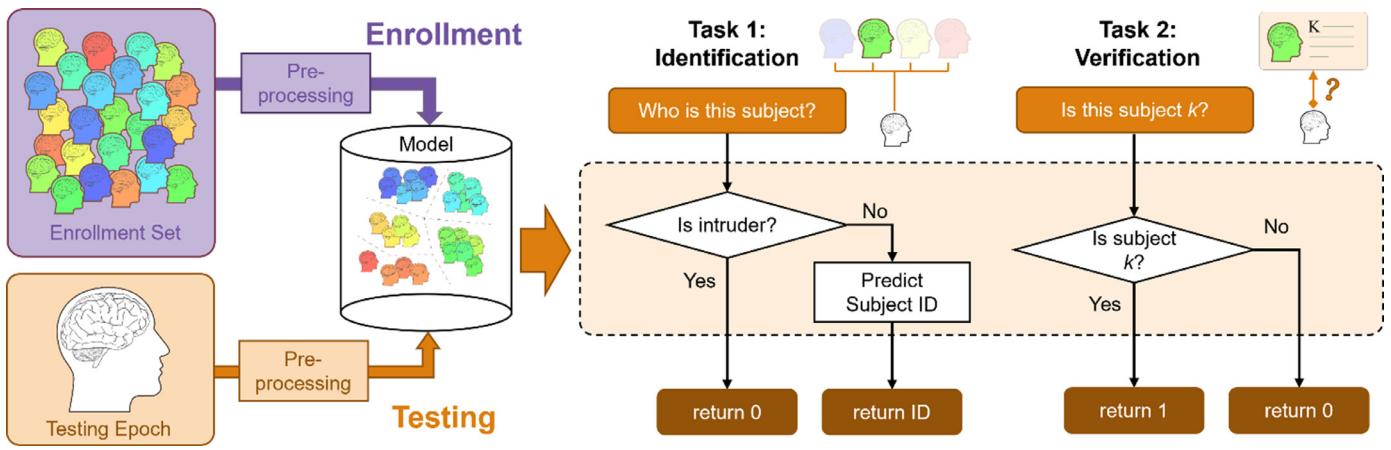
Based on the M³CV database, the EEG-based biometric competition was launched by focusing on the problems of personal identification and verification (Fig. 5). The contestant needs to train the classification model for the biometric system with the enrollment dataset. After the model has been trained, the contestant would be asked to complete the following two classification tasks: (1) Identification: Determine whether a given EEG signal comes from the enrollment set, and further determine

Table 4

Literature review on multi-session and multi-task EEG-based biometrics study.

Previous work	# Subjects	# Sessions	#Tasks	Feature	Classifier	Performance metric
La Rocca et al. (2013)	9	2	EC, EO	AR	LDC	ACC
Zeynali and Seyedarabi, 2019)	7	1-3	Rest, LT, MT, GFR, VC	PSD, AR, DWT, LEE, SE	SVM, NB	ACC, F1-score, AUC
Maiorana et al. (2016)	50	3	EC, EO	AR, PSD	L1, L2 & CS	ACC
Delpozo-Banos et al. (2018)	15	2	VEP, MT	PSD	NB, LDC	ACC, FAR, GAR
Maiorana, 2020)	45	5	EC, EO, MI, MT, VS, SI	AR, MFCC	CNNs & RNNs	Rank-1 IR
Arnaud-Gonzalez et al. (2021)	21	3	Rest, AS, MC, VC x, and VF x	AR, MFCC, PSD	HMM	ACC, AUC, EER
Kumar et al. (2021)	30	2-5	OB, FUW, IBA, MMI, PA, SSVEP, PAV	PSD	ix-vector	ACC, EER
M ³ CV (2022), [this work]	106	2	EC, EO, EP, SSVEP, SSAEP, SSSEP, P300, ME and SSVEP-SA	AR, MFCC, PSD	SVM, L2	Score

*Abbreviations: LT – Letter Task, MT – Math Task, GFR – Geometric Figure Rotation, DWT – Discrete Wavelet Transform, VC – Visual Counting, LEE – Log Energy Entropy, SE – Sample Entropy, NB – Naïve Bayes, LDC – Linear Discriminant Classifier, AUC – Area Under the Curve, FAR – False Acceptance Rate, GAR – Genuine Acceptance Rate, Rank-1 IR – Rank-1 Identification Rate, EER – Equal Error Rate, AR – Auto-Regressive, MFCC – Mel-Frequency Cepstrum Coefficients. VS – Visual Stimuli, HMM – Hidden Markov Models. Some abbreviations were omitted, which were already described in Table 1.

**Fig. 5.** The architecture of the EEG-based biometric system with the learning task of personal identification and verification.

the subject ID if the epoch is not judged as from an intruder and (2) Verification: Determine whether a given EEG signal comes from a certain subject.

3.3. Baseline model

3.3.1. Feature extraction

According to the literature review, three types of commonly used features were selected for EEG-based biometrics based on the EEG signal pre-processing:

- (1) Power Spectral Density (PSD): To estimate the power spectrum of each channel for each epoch, we used the Welch periodogram algorithm. Specifically, we divided the whole 4s epoch into 1s segments with 0.5s overlap, then we averaged the FFT of each segment using a hamming window. The PSD spectrum from 2 to 45 Hz was evenly divided into 12 frequency bands, then the band power of these frequency bands was extracted.
- (2) Mel Frequency Cepstral Coefficients (MFCCs): MFCCs are one of the most common features of speech recognition, and have been recently applied to the biometric analysis of EEG data (Arnaud-Gonzalez et al., 2021; Maiorana, 2020). MFCCs are extracted from the mel-filtered spectrum amplitudes s_m using log compression and discrete cosine transform, defined as

$$c_i = \sqrt{\frac{2}{M}} \sum_{m=1}^M \log(s_m) \cos\left(\frac{m - 0.5}{M}\pi i\right), \quad (1)$$

where L is the number of preserved cepstral coefficient, c_i is the i -th cepstral coefficient, with $i = 1, 2, 3, \dots, L$. M is the number of triangular

band-pass filters in a mel-scaled filter bank and L is the number of the first cepstral coefficients used to obtain signal representation. In this study, $M = 18$ and $L = 12$ were employed.

- (1) Autoregressive Coefficient (AR): AR features are commonly used for EEG-based identification and verification (Maiorana and Campisi, 2018). In this study, AR features were extracted from each epoch from a 12-th order autoregressive model created by solving the Yule-Walker equations. Thus 12 AR coefficients were extracted for each EEG channel.

3.3.2. Classification

In the biometric system, the verification task is a one-to-one classification problem and the identification task is a one-to-N classification problem. With the existence of the intruders, identification becomes a one-to-(N+1) classification problem. Considering the extensibility of the enrollment set, using N one-to-one weak classifiers from each subject to compose a one-to-N classifier is a practical way for the biometric system. The method can effectively avoid the retraining of the whole classifier after the enrolling of new subjects.

In this study, one-to-one classifiers, including the similarity-based method based on Euclidean distances (L2) and the One-class Support Vector Machine (SVM), were applied as the baseline classifiers for both verification and identification tasks in the subsequent analysis for simplicity. More specifically, after extracting features, an L2-based or SVM classification model was trained by epochs from each subject in the enrollment set.

In the verification task, the corresponding classifier would confirm the subject ID claimed by test epochs when the acquired score was greater than the rejection threshold δ . In an offline system, this thresh-

old was varied to obtain the equal error rate (*EER*), which was a commonly used threshold-independent measure in biometric verification tasks (Bidgoly et al., 2022), defined as the intersection point of false acceptance rate (*FAR*) curve and false rejection rate (*FRR*) curve. In an online system, the threshold has to be pre-determined, which was normally decided by obtaining *EER* on the training set (Marcel and Milian, 2007; Kang et al., 2018).

In the identification task, the intruder test was performed firstly, in which the test epoch was fed into each subject's classification model to generate a score, when the maximal score obtained from all subjects' models was below the threshold, the test epoch was judged as an intruder. For the L2-based model, the rule of judging the test epoch as an intruder was defined by Eq. (2)

$$\max_i \left\{ \frac{d(y_i, NN(y_i))}{d(z, y_i)} \right\} < \delta \quad (2)$$

where N is the number of subjects in the enrolled set, z is the test epoch, y_i is the nearest epoch of z from all epochs of i th subject with $i = 1, 2, 3, \dots, N$, $NN(y_i)$ is the nearest epoch of y_i from all epochs of i th subject, δ is the rejection threshold determined by *EER*. The subject ID would be assigned based on the maximum score from all subjects if the test epoch z passed the intruder detection. For an SVM-based classifier, the likelihood is applied instead of the distance in the L2-based classifier.

3.4. Performance metrics

To evaluate the performance of the different biometrics models, Accuracy (*Acc*), Equal Error Rate (*EER*), *Precision*, *Sensitivity*, *Specificity*, *F1-score*, T_{train} , T_{test} , and the final leaderboard *Score* were applied as performance metrics.

For offline evaluation, *Acc*, as the recognition accuracies, is the most commonly used for the identification task. For the verification task, *EER*, as the equal error rate of False Acceptance Rate (*FAR*) and False Rejection Rate (*FRR*), is the most commonly used (Maiorana, 2021b; Bidgoly et al., 2022; Kang et al., 2018). With the growth of the rejection threshold in the classifier model, *FAR* ($1 - \text{Sensitivity}$) increases and *FRR* ($1 - \text{Sensitivity}$) decreases, *EER* is defined as the intersection point of the *FAR* and *FRR* curve (Wu et al., 2018; Yingnan et al., 2019). Hence, for offline evaluation, *EER* is not only used to determine the rejection threshold during the training phase, but also used as the metric to evaluate the performance of the biometric algorithm.

For online evaluation, since *EER* is determined during the training phase, it could not be used as a performance metric for online testing. Hence, the most common and traditional performance metrics, *Acc*, *Precision*, *Sensitivity*, *Specificity*, *F1-score*, T_{train} , and T_{test} are applied for both the identification task (Fig. 6A) and the verification task (Fig. 6B) (Wang et al., 2019; Kong et al., 2019; Abolfazl et al., 2019). It should be noticed the definition of *Acc*, *Precision*, *Sensitivity*, *Specificity*, and *F1-score* would be different between the multi-class identification task in Fig. 6A and two-class verification task in Fig. 6B. T_{train} and T_{test} represent the time of training and testing. In this study, T_{train} represents the time of training the classifier by the epochs from Sub11-Sub18 separately. T_{test} represents the time of testing all epochs of Sub11-Sub20. All the test are run on the computer with CPU i9-9900KF Inter(R) Core(TM).

To avoid information leakage, the performance metrics of *Acc*, *Precision*, *Sensitivity*, *Specificity*, and *F1-score* are only calculated in the mock online test. For the real online competition, the evaluation of these metrics needs the true labels from the testing set. Hence, only the leaderboard score, as the weighted sum of *Acc* in the identification and verification task (Acc_I and Acc_V) in Fig. 6C, was reported for online submission, which is provided by the platform of Kaggle. The leaderboard score was used as the unique ranking in the competition to evaluate the contestant's overall performance in both tasks to decide the final winner.

3.5. Online competition

3.5.1. Dataset composition

As is shown in Fig. 7, the whole set for this competition was divided into three parts: Enrollment Set, Calibration Set, and Testing set (containing 11 intruders).

- Enrollment Set: it provides the 1st session of EEG from 95 subjects. The enrollment set is used for the contestants to train the biometric model.
- Calibration set: the EEG data of the 2nd session from 20 subjects is provided in the calibration set, which is used to help the contestants get familiar with the data format, refine effective feature extraction, and tune hyper-parameters of the machine learning model. The number of subjects in the calibration set is relatively small as compared with the enrollment set and testing set.
- Testing set: it is used to evaluate the performance of the algorithm in the competition with the public and private leaderboards, for which epochs belonging to public or private leaderboards are hidden. The testing epochs come from the 2nd session of 86 subjects, among which 11 subjects are treated as an intruder.

3.5.2. Organization of the competition

The challenge has already been launched currently and would be closed on Apr 30th, 2023. Challengers were ranked based on the score of recognition accuracy of their submission computed on the private leaderboard. The platform Kaggle is used to hold the competition. During the challenge, contestants can submit their solutions on the Kaggle, and we provided the contestants with the score computed on the public leaderboard. At the end of the challenge, contestants were ranked based on the score of these submissions computed on the private leaderboard. Later submission is still allowed on the Kaggle platform for the researcher to evaluate their methods.

3.5.3. Competition platform

Kaggle is an online community platform for data scientists and machine learning enthusiasts, which provides the world's largest data science community with powerful tools and resources. The link to the platform for our competition are as follows,

- Kaggle: <https://www.kaggle.com/competitions/eeg-biometric-competition>

3.5.4. Data and code availability

The public dataset with the code of the baseline methods and an example submission file are available at

- Kaggle: <https://www.kaggle.com/competitions/eeg-biometric-competition/data>.

3.6. Benchmarks

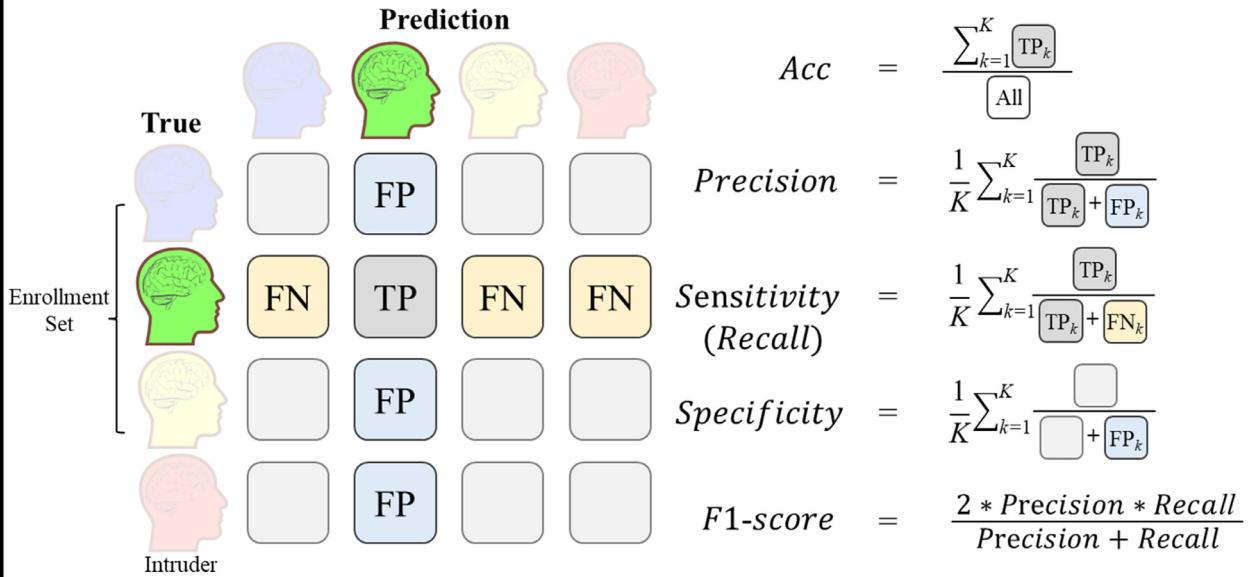
To have a comprehensive understanding of the challenges in the EEG-based biometrics competition, we divided the benchmark test into three levels.

- (1) **Offline evaluation** aims to reveal the challenge of the competition with the increasing population size and cross-session, and cross-task;
- (2) **Mock online test** aims to evaluate the performances of different baseline models without information leakage.
- (3) **Online submission** provides an example code and the baseline score in the public.

3.6.1. Offline evaluation

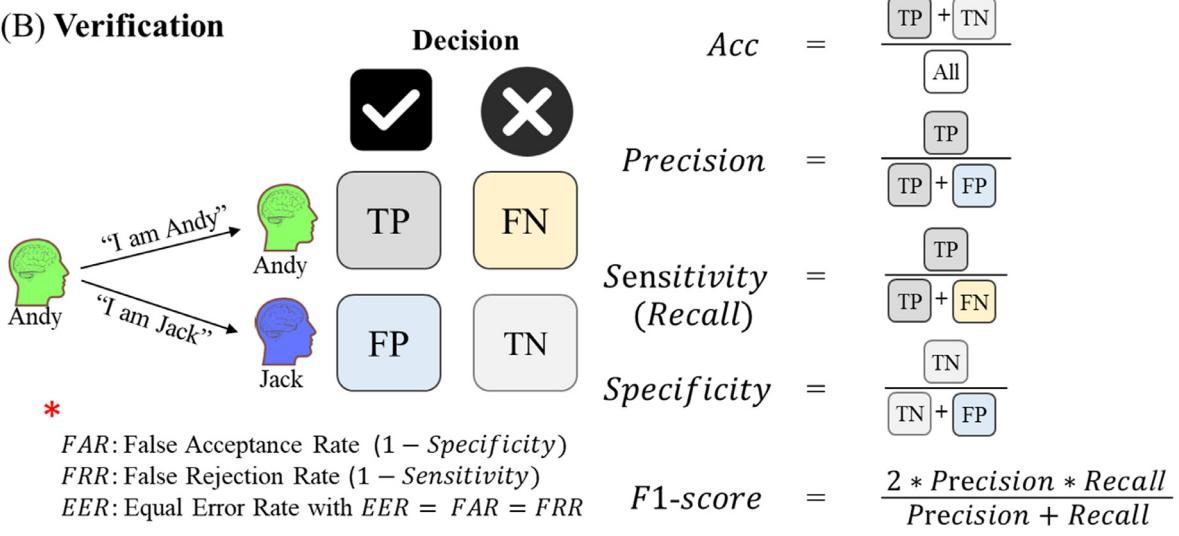
For offline evaluation, three analyses were performed to show the challenge of the increasing population sizes, cross-session, and cross-task tests in this EEG-based biometrics competition. These analyses were

(A) Identification



* K is the number of subjects in enrollment set plus 1, one for the class of intruder.

(B) Verification



(C) Leaderboard Metric

$$\text{Score} = \frac{N_I}{N_I + N_V} * \text{Acc}_I + \frac{N_V}{N_I + N_V} * \text{Acc}_V$$

* N_I : Epoch numbers for identification tasks
 N_V : Epoch numbers for verification tasks

Acc_I : The accuracy for identification tasks

Acc_V : The accuracy for verification tasks

performed on the data from 20 subjects in the calibration set and enrollment set, whose subject ID of the second session was known. To avoid any possible information leakage, all the analyses were limited to these 20 subjects, not any data and label information from the testing set has been used. The intruder was not considered. The train/test pipeline and performance metric for these analyses were detailed as follows:

- (1) **Testing on increasing population sizes:** To evaluate the influence of increasing population sizes on the performance of EEG-based biometrics, we performed a within-task and cross-session testing the number of subjects was increased from 2 to 20 with the step of 2 subjects. To avoid training bias, subjects were randomly selected 10 times for each number of subjects. The epochs

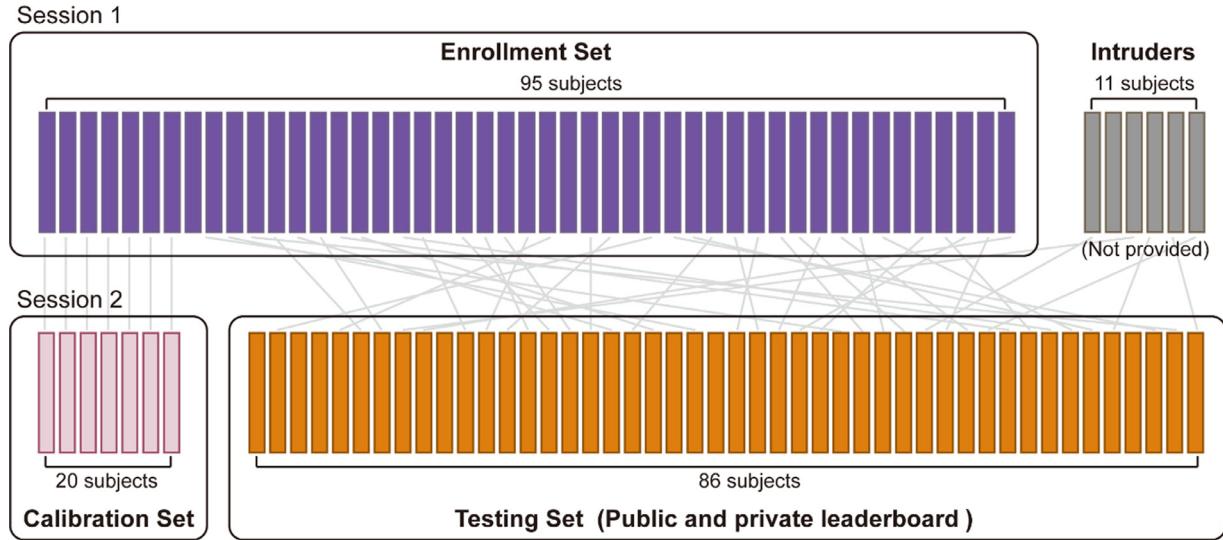


Fig. 7. The segmentation of M³CV for online competition. The whole set was divided into three parts: Enrollment Set, Calibration Set, and Testing set (containing 11 intruders).

from different EEG tasks were used for the testing separately. For each task, the data from the first session was used to train, and the data from the second session was used to test.

- (2) **Testing across tasks:** To compare the performance between within and cross tasks, we performed cross-session testing with a population size of 20. EEG data of a particular task from one session (session 1/session 2) was used to train the model, and EEG data of another task from the other session (session 2/session 1) was used to test. And then the training set and the testing set would be exchanged. Since AEP, SEP and VEP are in the same runs, we treated them as one condition/task, EP (i.e., Evoked Potential). Also, the term ME (Motor execution) was used to represent the LH, RH, and RF conditions.
- (3) **Testing across sessions:** To compare the performance between within and cross sessions, we performed the within-task test with a population size of 20. To make a fair comparison of the performance between within-session and cross-session tests, the training set was the first half of the epochs in the first session, which was the same for the within-session tests and cross-session tests. For the within-session test, the testing set was the second half of the epochs in the first session. For the cross-session tests, the testing set was all epochs from the second session.

In the offline evaluation, *Acc* was adopted as the performance metric of the identification task, which is the larger the better, and *EER* was used for the verification task, which is the smaller the better (Maiorana, 2021b; Bidgoly et al., 2022; Kang et al., 2018).

3.6.2. Mock online test

To evaluate the performances of different models without any information leakage from the testing set, mock online test was performed on the 20 subjects in the true calibration set and enrollment set. Similar to the dataset composition of the online competition, we divided these 20 subjects into the mock enrollment set, calibration set, and testing set.

- The mock enrollment training set consisted of epochs from the first sessions of the first 18 subjects (Sub001 - Sub018).
- The mock calibration set consisted of epochs from the first sessions of the first 10 subjects (Sub001 - Sub010).
- The mock testing set consisted of epochs from the second session of another 10 subjects (Sub011- Sub020), in which Sub019 and Sub020 were set as intruders. The epochs from the first session of Sub019 and Sub020 were not used.

For mock online test, the performance metrics, *Accuracy* (*Acc*), *Precision*, *Sensitivity*, *Specificity*, *F1-score*. T_{train} , T_{test} and the final *Score* were applied for both identification and verification tasks. As is shown in Fig. 6, *Acc*, *Precision*, *Sensitivity*, *Specificity*, and *F1-score* were adopted to comprehensively evaluate the prediction result, which is the larger the better. T_{train} and T_{test} were the evaluation of the time consumption for training and testing, which is the smaller the better. More specifically, T_{train} represents the training time of the classifier by the epochs from Sub11-Sub18 in the first session. T_{test} represents the testing for all epochs of Sub11-Sub20 in the second session. Finally, the *Score* is the overall accuracy considering both tasks, which decides the ranking of different baseline models in the mock online test, the larger the better.

3.6.3. Online submission

For online submission, an example code was provided with different extracted features (PSD, AR, MFCC) and classification (L2 and SVM), in which the optimal threshold was determined by making the $FAR = FRR$ on the 20 subjects in the calibration set and enrollment set, whose subject ID of two sessions was known. An example code and the submission file with the method of PSD+L2 with were provided on the website of Kaggle (<https://www.kaggle.com/competitions/eeg-biometric-competition/data>).

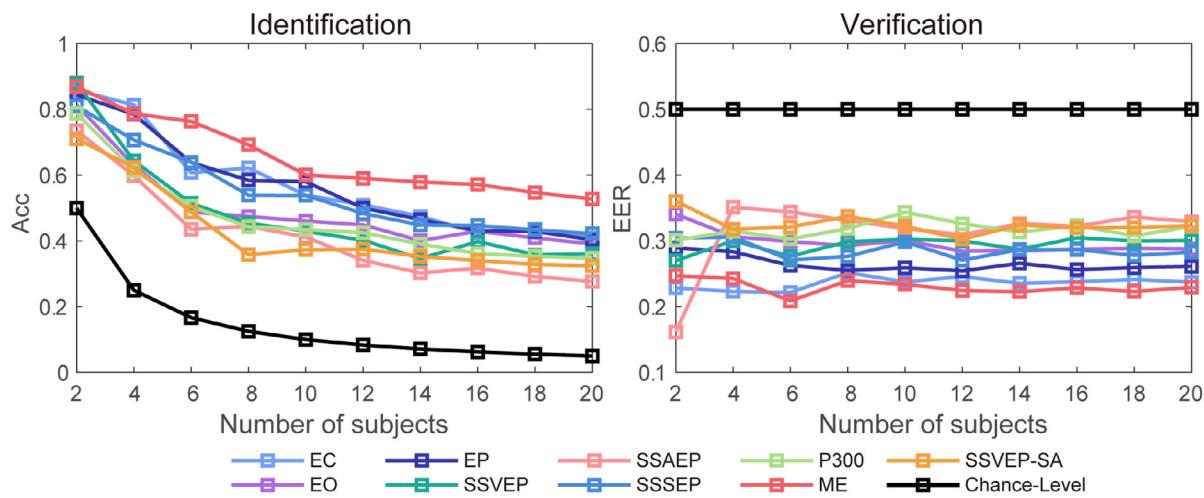
To avoid the information leakage of the subject ID for the epochs in the testing set, only the leaderboard *Score* of the total recognition accuracy was used to evaluate the performance, which can be obtained from the platform of Kaggle. We did not set multiple leaderboards for the competition. As a unified and comprehensive index will make the competition focused and competitive, the leaderboard *Score* was used to evaluate the contestant's overall performance in both identification and verification tasks to decide the final winner.

4. Benchmark results

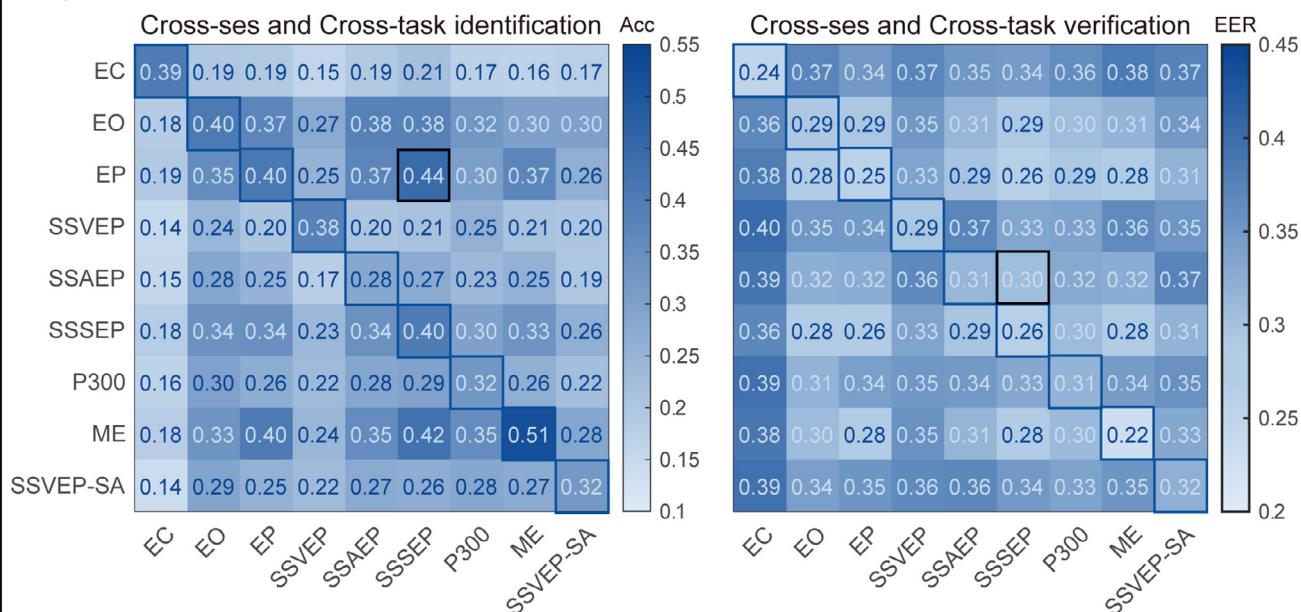
4.1. Offline evaluation

Since this section only aims to evaluate the performance degradation resulting from larger population sizes, cross-session, and cross-task tests, only the MFCC feature and SVM classifier was adopted. The results for the investigation of the uniqueness of individuals, robustness to mental state changes, and permanence over recording sessions of EEG-based biometrics are shown in Fig. 8(A–C) correspondingly.

A) Increasing population sizes



B) Within-task versus Cross-task



C) Within-session versus Cross-session

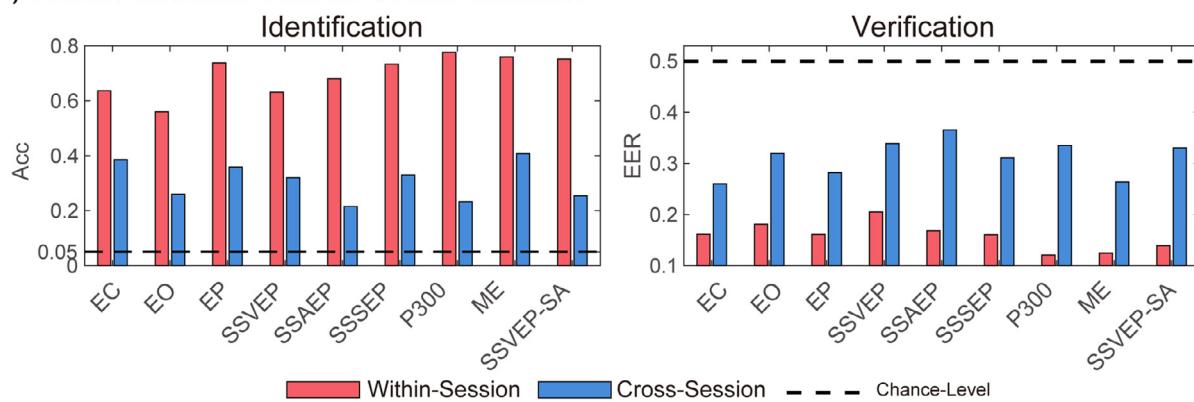


Fig. 8. The influence of (A) increasing population sizes, (B) cross-task tests, and (C) cross-session tests on the performance degradation of EEG-based biometric systems. *Acc* is used to evaluate the performance in the identification task. The larger value of *Acc* indicates a better result. *EER*, as the equal error rate, is used to evaluate the performance in the verification task. The smaller value of *EER* indicates a better result. Chance levels in Fig. 8A and C are marked by black curves. The influence of increasing population sizes (A) and cross-task tests (B) were both evaluated in the data from the second session. These results were obtained on 20 subjects in the enrollment set and calibration set. Left: identification results. Right: verification results.

Table 5

Model evaluation for the combination of different features (PSD, MFCC, AR) and different classifiers (L2 and SVM) on the mock online test.

Performance Metric		SVM			L2		
		PSD	MFCC	AR	PSD	MFCC	AR
Identification	<i>Acc</i>	0.461	0.437	0.284	0.531	0.410	0.373
	<i>Precision</i>	0.458	0.439	0.346	0.544	0.416	0.404
	<i>Sensitivity</i>	0.596	0.670	0.589	0.570	0.427	0.412
	<i>Specificity</i>	0.933	0.930	0.910	0.941	0.925	0.924
	<i>F1-score</i>	0.517	0.530	0.436	0.557	0.421	0.408
Verification	<i>Acc</i>	0.743	0.721	0.474	0.746	0.757	0.438
	<i>Precision</i>	0.235	0.219	0.136	0.256	0.242	0.138
	<i>Sensitivity</i>	0.694	0.695	0.797	0.811	0.668	0.880
	<i>Specificity</i>	0.749	0.724	0.438	0.738	0.767	0.389
	<i>F1-score</i>	0.351	0.333	0.233	0.389	0.355	0.239
Computation Time	<i>T_{train}</i>	1.70s	1.36s	1.25s	0.75s	0.69s	0.71s
	<i>T_{test}</i>	1.46s	1.42s	1.40s	3.28s	3.22s	3.29s
	<i>Score</i>	0.712	0.690	0.453	0.721	0.718	0.431

* PSD: Power Spectral Density, MFCCs: Mel Frequency Cepstral Coefficients, AR: Autoregressive Coefficient, L2: Euclidean distances, and SVM: Support Vector Machine. *Acc*, *Precision*, *Sensitivity*, *Specificity*, *F1-score* were applied to evaluate the recognition efficiency for both identification and verification task which is the larger the better. *T_{train}* and *T_{test}* were the time consumption for training and testing., which is the smaller the better. The leaderboard metric *Score* is the overall accuracy considering both tasks, which is the larger the better. The numbers in bold indicate the best results.

In terms of the uniqueness of individuals, it can be observed that the accuracy decreased for each task as the population sizes increased from 2 to 20. The averaged recognition accuracy decreased from 0.81 to 0.34. It should be noted that with the different population sizes, the chance levels are different. With the population sizes rising to a larger scale (e.g., 106 subjects in our testing set), the identification problem for machine learning algorithms to identify one subject would be more challenging. Since the verification task is a one-to-one matching problem, increasing the population size will not lead to its performance degradation.

Fig. 8B showed the robustness of EEG-based biometrics to mental state changes, in which the row represents the task used for training, and the column represents the task used for testing. The blue square border on the diagonal elements in Fig. 8B indicates the result of the within-task test, as compared with the result of the cross-task test in the off-diagonal elements. In the left figure of Fig. 8B, the black square border on the off-diagonal elements (where EP was used to train and SSSEP was used to test) shows a higher classification accuracy than the blue square border (where EP was used to train and test) in the third row, which is 0.44 versus 0.40. There was a similar case for the verification task, when SSAEP was used to train, the best result was obtained when SSSEP was used to test. Except for these two, the best performance for other rows was achieved at the diagonal line, in which the training task and the testing task is the same. This result is consistent for both identification and verification tasks. Due to the unequal number of training samples for each task, it's unfair to compare the error rate across different rows. The within-task and cross-task test results were obtained in cross-session settings.

As for the permanence across recording sessions in Fig. 8C, the averaged *EER* increased from 0.16 to 0.31, and the average accuracy for identification decreased from 0.70 to 0.31.

4.2. Mock online test

To make a comprehensive comparison of different methods without information leakage, we hold the mock online test by using the first 20 subjects in the enrollment set and calibration set to mock the online biometric competition. Table 5 shows the model performances on the mock online test for the combination of extracted features (PSD, MFCC, AR) and classification (L2 and SVM) methods. The value of *Score*, which decides the ranking on the mock leaderboard for six models, is also pro-

vided. PSD+L2 provided the best result as *Score* = 0.721. It can be observed that the precision for the verification task is quite low, which is caused by the class imbalance in the mock online test.

In terms of extracted features, the AR features have the worst performance no matter which classifier was used. PSD features have a better or similar performance, as compared with more elaborated features such as MFCC across each metric. It is interesting to note that with the threshold determined by the equal error rate (i.e., $FAR = FRR$) in an independent training set, the performance of MFCC and AR features are very different on the metrics of specificity ($1 - FAR$) and sensitivity ($1 - FRR$). Especially, the specificity of the AR feature is 0.438 and 0.389 for L2 and SVM classifiers, but the corresponding sensitivity is 0.797 and 0.880, exhibiting strong instability for AR-based verification systems.

In terms of classifier, the SVM classifier performed a little better than L2 classifier with the PSD and MFCC features. As for the time consumption, since the L2-based classifier has to compare the new input with all points in the training set, its evaluation time *T_{test}* is slower than the SVM classifier. Since the training and testing time depend on the code, the hardware, and the software environment, it's difficult to make a relatively consistent evaluation for online competition.

4.3. Online submission

For the example submission, the PSD+L2 method obtained a *Score* of 0.4418 in public leaderboard. It should be noted that the method with a better result on the mock online test may not necessarily directly lead to higher scores in the online submission. That is caused by the different samples used for training and testing. For the same reason, it is also possible that the score varies from public leaderboard to private leaderboard.

5. Discussion and conclusion

5.1. Challenges in EEG-based biometrics competition

In the field of EEG-based biometrics, many studies have mentioned that their major limitation is lacking a database containing large population sizes, recorded under different conditions and different sessions (Wang et al., 2020; Chan et al., 2018; Delpozo-Banos et al., 2018). Based on our proposed M³CV database, we launch the EEG-based biometric

competition. In view of the current development of EEG-based biometric technology, this competition proposes the following challenges for machine learning algorithms from many aspects:

1. **Challenge with large population size:** the population size of the testing database is an important indicator of whether the current EEG-based biometric technology can be practically applied. As illustrated in Fig. 8(A), the identification accuracies decreased greatly with the increase of the population size. The identification task on the online competition with a population size of 106 would be more challenging for the machine learning algorithm.
2. **Challenge for cross-session tests:** a practical biometric system should be robust to cross-session variability. However, as illustrated in Fig. 8(C), the machine learning model would easily achieve a good performance in the within-session test but degraded severely in the cross-session test. Hence, the cross-session test would be one big challenge in this competition.
3. **Challenge for cross-task tests:** similar to the cross-session tests, the robustness of the biometric system to the cross-task variability is also a challenge in this competition. For real application, the biometric system should recognize the subject even when they are in different mental states.
4. **Challenge for intruders:** Whether the biometric system can prevent an attack from intruders outside the enrollment set is determined by the identification performance on an open set (Kumar et al., 2021). That would lead to very different strategies for the classifier design in terms of machine learning techniques (Gunther et al., 2017).

To our best knowledge, no previous study has tested the permanence and robustness of EEG-based biometric features over 100 subjects in an open-set setting. Hence, the M³CV database can facilitate the development of advanced machine learning algorithms for EEG-based biometrics.

It should be mentioned that there are samples from the same subject in both the testing set and calibration set, which may lead to these epochs in the testing set being easy to recognize. However, we chose this scheme to construct the testing set and calibration set for two reasons. First, the large sample size is one of the major challenges for this competition. We had to divide the dataset for calibration, which includes 20 subjects in the competition. But if we did not use this part of subjects at all in the testing set, the number of subjects would be decreased by almost 20%, which makes the competition less challenging. Second, this type of information leakage may also occur in practical applications. Once the biometric model has been trained, users can use the biometric system immediately. For using the system, it is not necessary for them to re-wear the EEG cap.

5.2. From competition to real application

Due to the limitations of the competition conditions, there may still be some distance away from the real application.

Firstly, the competition needs to have a unique ranking. Hence the *score* was applied to comprehensively evaluate the contestant's performance in the identification and verification task to decide the final winner. However, the evaluation of a system's performance should consider multiple aspects. On the calibration set, accuracy, precision, sensitivity, specificity, F1-score, T_{train} , and T_{test} have been proposed as performance metrics for a comprehensive evaluation of the system. For the submission of each team, the error rate under different conditions can be calculated offline, but the computation complexity of training and testing is difficult to make a relatively consistent evaluation. Hence, it is required for the contestants to report the time for training and testing, and their hardware and software environment.

Furthermore, the baseline models provided in the studies are all one-class classifiers (L2 and one-class SVM). If a new subject is added to the enrollment set, the biometrics system just needs to add the model for

the new coming subject alone. It doesn't need to retrain the whole classifier model. We believe the multi-class classifier would achieve better performance in the competition. But one-class classifier would be more practical in the real application.

5.3. Intra- and inter-subject variability

Intra- and inter-subject variability poses a great challenge for the interpretation and decoding of EEG signals (Seghier and Price, 2018; Wei et al., 2021). Traditional ERP analysis and statistical test methods were typically used to analyze the group-level commonality of EEG signals, while machine learning techniques are considered to be more powerful in dealing with intra- and inter-subject variability. The dataset for this competition is not only restricted to the study of EEG-based biometric algorithms but can also be used for other applications, such as cross-session reliability analysis of EEG signals from different tasks, transfer learning in cross-subject, and cross-session BCI studies. Furthermore, the cross-run and cross-trial variability within subject is also a meaningful topic to study. The multi-run EEG recording, which included resting-state, transient-state sensory, and motor execution, made the M³CV database useful in studies mentioned above.

5.4. Advanced machine learning methods

To hold the competition, we only provide a few simplistic models as the baseline to help the contestant to calibrate their work, no advanced machine learning approaches have been applied in this study. Nevertheless, different strategies have been proposed to deal with intra-subject variability caused by mental state changes and the influence of external factors, such as varying electrode placement and electrical impedance. For example, some studies have started to explore the evidence of task-independent person-specific signatures in EEG. Kong et al. (2019) have claimed that phase synchronization of EEG signals has task-free biometric properties but it lacked cross-session evaluation, while Valizadeh et al. (2019) conducted a cross-session evaluation for various connectivity measures, only 5 participants were measured twice. Besides, some advanced machine learning methods, such as low-rank learning (Kong et al., 2018) and adversarial deep learning (Ozdenizci et al., 2019) have also been applied to deal with mental state changes. Also, some strategies, such as multi-task learning (Sun, 2008) and instance-based learning (Yang et al., 2022) have been proposed to obtain better cross-session results.

Undoubtedly, deep learning is the biggest advance in machine learning in recent years, which has also been applied in the EEG-based biometrics study (Behrouzi and Hatzinakos, 2022; Jin et al., 2021; Wang et al., 2019; Maiorana, 2021b). However, due to the lack of support from large datasets, the generalizability of deep learning has always been doubtful in this field. On the other hand, most of the newly proposed ideas of transfer learning are based on the framework of deep learning. Based on this competition on the M³CV database, it is expected that some cross-session and cross-task deep learning methods can emerge. Furthermore, it is also expected that these methods can not only work well in the field of EEG-based biometrics but also provide a universal solution in dealing with Intra- and inter-subject variability of the EEG signals.

6. Conclusion

The lack of a large-scale comprehensive EEG database, which contains data recorded from multiple subjects in multiple sessions and over multiple tasks, limits our understanding of intra- and inter-subject variability and hinders the development of advanced machine learning algorithms. In this study, we established an M³CV database with 106 subjects, 2 sessions, and 6 tasks to reveal the commonality and variability of cross-subject, cross-session, and cross-task EEG signals. Based on this

M³CV dataset, a machine learning competition about EEG-based biometrics was launched to help this field grow. Except for advancing the development of machine learning algorithms for EEG decoding (such as biometrics and BCI), we believe the M³CV database can help researchers gain a deep understanding of the relationship between different types of EEG signals, such as resting state, motor, sensory-related, or cognitive-related EEG signals.

Data and code availability statement

This is an open-access database. All the data for Enrollment Set, Calibration Set and Testing Set with the baseline methods and an example submission file can be download.

- Kaggle: <https://www.kaggle.com/competitions/eeg-biometric-competition/data>.

The detail description about the EEG-based biometric competition can be referred to

- Kaggle: <https://www.kaggle.com/competitions/eeg-biometric-competition>.

Credit authorship contribution statement

Gan Huang: Conceptualization, Formal analysis, Methodology, Writing – original draft. **Zhenxing Hu:** Conceptualization, Formal analysis, Methodology, Writing – original draft. **Weize Chen:** Methodology, Formal analysis. **Shaorong Zhang:** Writing – review & editing. **Zhen Liang:** Conceptualization, Formal analysis. **Linling Li:** Data curation. **Li Zhang:** Formal analysis. **Zhiguo Zhang:** Supervision, Conceptualization, Writing – review & editing.

Acknowledgments

We would like to thank the reviewers who contributed their time to review manuscripts for the three rounds of revision. We thank Hu Li and Kong Yazhuo from the Institute of Psychology, Chinese Academy of Sciences, Peng Weiwei from the School of Psychology, Shenzhen University, Liang Meng from the School of Medical Imaging, Tianjin Medical University, Liu Jixin from the School of Life Science and Technology, Xidian University for their suggestions on our experimental design.

This work was supported by the National Natural Science Foundation of China (No. 62271326, 81871443, 61906122, and 81901831), the Science, Technology, and Innovation Commission of Shenzhen Municipality Technology Fund (No. JCYJ20190808173819182), the Shenzhen Science and Technology Program (No. JSGG20210713091811038), the Shenzhen Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (No. 2022SHIBS0003), the Shenzhen's Sanming Project of Medicine (No. SZSM202111009), the Shenzhen Special Project for Sustainable Development (KCXFZ20201221173400001).

References

- Abolfazl, S., Riener, R., Elmer, S., Lutz, J., 2019. Decrypting the electrophysiological individuality of the human brain : identification of individuals based on resting-state EEG activity. *Neuroimage* 197, 470–481. doi:[10.1016/j.neuroimage.2019.04.005](https://doi.org/10.1016/j.neuroimage.2019.04.005).
- Alotaiby, T.N., Alshebeili, S.A., Alshawi, T., Ahmad, I., Abd El-Samie, F.E., 2014. EEG seizure detection and prediction algorithms: a survey. *EURASIP J. Adv. Signal Process.* doi:[10.1186/1687-6180-2014-183](https://doi.org/10.1186/1687-6180-2014-183).
- Arnao-Gonzalez, P., Katsigiannis, S., Arevalillo-Herraez, M., Ramzan, N., 2021. BED: a new data set for EEG-based biometrics. *IEEE Internet Things J.* 8, 12219–12230. doi:[10.1109/JIOT.2021.3061727](https://doi.org/10.1109/JIOT.2021.3061727).
- Autthasan, P., Chaisaen, R., Sudhawiyangkul, T., Rangpong, P., Kiatthaveephong, S., Dilokthanakul, N., Bhakdisongkhram, G., Phan, H., Guan, C., Wilairapsitporn, T., 2022. MIN2Net: end-to-end multi-task learning for subject-independent motor imagery EEG classification. *IEEE Trans. Biomed. Eng.* 69, 2105–2118. doi:[10.1109/TBME.2021.3137184](https://doi.org/10.1109/TBME.2021.3137184).
- Behrouzi, T., Hatzinakos, D., 2022. Graph variational auto-encoder for deriving EEG-based graph embedding. *Pattern Recognit.* 121, 108202. doi:[10.1016/j.patcog.2021.108202](https://doi.org/10.1016/j.patcog.2021.108202).
- Berger, H., 1929. Über das elektrenkephalogramm des menschen. *Arch. Psychiatr. Nervenkr.* 87, 527–570. doi:[10.1007/BF01797193](https://doi.org/10.1007/BF01797193).
- Bidgoly, A.J., Bidgoly, H.J., Arezoumand, Z., 2022. Towards a universal and privacy preserving EEG-based authentication system. *Sci. Rep.* 12, 1–12. doi:[10.1038/s41598-022-06527-7](https://doi.org/10.1038/s41598-022-06527-7).
- Boshra, R., Dhindsa, K., Boursalé, O., Ruiter, K.I., Sonnadar, R., Samavi, R., Doyle, T.E., Reilly, J.P., Connolly, J.F., 2019. From group-level statistics to single-subject prediction: machine learning detection of concussion in retired athletes. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 1492–1501. doi:[10.1109/TNSRE.2019.2922553](https://doi.org/10.1109/TNSRE.2019.2922553).
- Brunner, C., Leeb, R., Müller-Putz, G., Schlägl, A. and Pfurtscheller, G., 2008. BCI Competition 2008—Graz data set A. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16, pp. 1–6.
- Cahn, B.R., Polich, J., 2006. Meditation states and traits: EEG, ERP, and neuroimaging studies. *Psychol. Bull.* 132, 180–211. doi:[10.1037/0033-295X.132.2.180](https://doi.org/10.1037/0033-295X.132.2.180).
- Chan, H.L., Kuo, P.C., Cheng, C.Y., Chen, Y.S., 2018. Challenges and Future Perspectives on Electroencephalogram-Based Biometrics in Person Recognition. *Front. Neuroinform.* 12, 1–15. doi:[10.3389/fninf.2018.00066](https://doi.org/10.3389/fninf.2018.00066).
- Chen, Y., Atnafu Ayaleh, D., Wendimagegn, T., Roh Kakao corp, M.-C., Dessalegn Atnafu, A., Schlattner, I., Tariku Weldesadik, W., Roh, M.C., Joong Kim, H., Lee, S.-W., Blankertz, B., Fazli, S., 2016. A high-security EEG-based login system with RSVP stimuli and dry electrodes. *ieeexplore.ieee.org* 11. doi:[10.1109/TIFS.2016.2577551](https://doi.org/10.1109/TIFS.2016.2577551).
- Cho, H., Ahn, M., Ahn, S., Kwon, M., Jun, S.C., 2017. EEG datasets for motor imagery brain-computer interface. *Gigascience* 6. doi:[10.1093/gigascience/gix034](https://doi.org/10.1093/gigascience/gix034).
- Craik, A., He, Y., Contreras-Vidal, J.L., 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *J. Neural Eng.* 16. doi:[10.1088/1741-2552/AB0AB5](https://doi.org/10.1088/1741-2552/AB0AB5).
- Debie, E., Moustafa, N., Vasilakos, A., 2021. Session invariant EEG signatures using elicitation protocol fusion and convolutional neural network. *IEEE Trans. Dependable Secur. Comput.* doi:[10.1109/TDSC.2021.3060775](https://doi.org/10.1109/TDSC.2021.3060775).
- Delpozzo-Banos, M., Travieso, C.M., Alonso, J.B., John, A., 2018. Evidence of a task-independent neural signature in the spectral shape of the electroencephalogram. *Int. J. Neural Syst.* 28, 1–20. doi:[10.1142/S0129065717500356](https://doi.org/10.1142/S0129065717500356).
- Dietrich, A., Kanso, R., 2010. A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol. Bull.* 136, 822–848. doi:[10.1037/a0019749](https://doi.org/10.1037/a0019749).
- Fisher, A.J., Medaglia, J.D., Jeronimus, B.F., 2018. Lack of group-to-individual generalizability is a threat to human subjects research. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6106–E6115. doi:[10.1073/pnas.1711978115](https://doi.org/10.1073/pnas.1711978115).
- Fröhner, J.H., Teckentrup, V., Smolka, M.N., Kroemer, N.B., 2019. Addressing the reliability fallacy in fMRI: Similar group effects may arise from unreliable individual effects. *Neuroimage* 195, 174–189. doi:[10.1016/j.neuroimage.2019.03.053](https://doi.org/10.1016/j.neuroimage.2019.03.053).
- Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U. S. A.* 78, 2643–2647. doi:[10.1073/pnas.78.4.2643](https://doi.org/10.1073/pnas.78.4.2643).
- Gaspar, C.M., Rousselet, G.A., Pernet, C.R., 2011. Reliability of ERP and single-trial analyses. *Neuroimage* 58, 620–629. doi:[10.1016/j.neuroimage.2011.06.052](https://doi.org/10.1016/j.neuroimage.2011.06.052).
- Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Miller, J.E., Moody, G.B., Peng, C.K., Stanley, H.E., 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101. doi:[10.1161/01.cir.101.23.e215](https://doi.org/10.1161/01.cir.101.23.e215).
- Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glärtner, T., Halchenko, Y.O., Handwerker, D.A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B.N., Nichols, T.E., Pellman, J., Poline, J.B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J.A., Varoquaux, G., Poldrack, R.A., 2016. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. data* 3. doi:[10.1038/sdata.2016.44](https://doi.org/10.1038/sdata.2016.44).
- Gui, Q., Jin, Z., Xu, W., 2015. Exploring EEG-based biometrics for user identification and authentication. In: Proceedings of the 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) doi:[10.1109/SPMB.2014.7002950](https://doi.org/10.1109/SPMB.2014.7002950).
- Gunther, M., Cruz, S., Rudd, E.M., Boult, T.E., 2017. Toward open-set face recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2017-July, pp. 573–582. doi:[10.1109/CVPRW.2017.85](https://doi.org/10.1109/CVPRW.2017.85).
- Herrmann, C.S., 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Exp. Brain Res.* 137, 346–353. doi:[10.1007/s002210100682](https://doi.org/10.1007/s002210100682).
- Hu, Z., Zhang, Z., Liang, Z., Zhang, L., Li, L., Huang, G., 2022. A new perspective on individual reliability beyond group effect for event-related potentials: a multisensory investigation and computational modeling. *Neuroimage* 250, 118937. doi:[10.1016/j.neuroimage.2022.118937](https://doi.org/10.1016/j.neuroimage.2022.118937).
- Huang, G., Hu, Z., Zhang, L., Li, L., Liang, Z., Zhang, Z., 2020. Removal of eye-blinking artifacts by ICA in cross-modal long-term EEG recording. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). EMBS 2020-July, pp. 217–220. doi:[10.1109/EMBC44109.2020.9176111](https://doi.org/10.1109/EMBC44109.2020.9176111).
- Infantolino, Z.P., Luking, K.R., Sauder, C.L., Curtin, J.J., Hajcak, G., 2018. Robust is not necessarily reliable: from within-subjects fMRI contrasts to between-subjects comparisons. *Neuroimage* 173, 146–152. doi:[10.1016/j.neuroimage.2018.02.024](https://doi.org/10.1016/j.neuroimage.2018.02.024).
- Jeong, J.H., Cho, J.H., Shim, K.H., Kwon, B.H., Lee, B.H., Lee, D.Y., Lee, D.H., Lee, S.W., 2020. Multimodal signal dataset for 11 intuitive movement tasks from single upper extremity during multiple recording sessions. *Gigascience* 9, 1–15. doi:[10.1093/GIGASCIENCE/GIA098](https://doi.org/10.1093/GIGASCIENCE/GIA098).
- Jin, X., Tang, J., Kong, X., Peng, Y., Cao, J., Zhao, Q., Kong, W., 2021. CTNN: a convolutional tensor-train neural network for multi-task brainprint recognition. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 103–112. doi:[10.1109/TNSRE.2020.3035786](https://doi.org/10.1109/TNSRE.2020.3035786).
- Kang, J.H., Jo, Y.C., Kim, S.P., 2018. Electroencephalographic feature evaluation for

- improving personal authentication performance. *Neurocomputing* 287, 93–101. doi:[10.1016/j.neucom.2018.01.074](https://doi.org/10.1016/j.neucom.2018.01.074).
- Kappenman, E.S., Farrens, J.L., Zhang, W., Stewart, A.X., Luck, S.J., 2021. ERP CORE: An open resource for human event-related potential research. *Neuroimage* 225, 117465. doi:[10.1016/j.neuroimage.2020.117465](https://doi.org/10.1016/j.neuroimage.2020.117465).
- Koike-Akino, T., Mahajan, R., Marks, T.K., Wang, Y., Watanabe, S., Tuzel, O., Orlik, P., 2016. High-accuracy user identification using EEG biometrics. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS) 2016-October, pp. 854–858. doi:[10.1109/EMBC.2016.7590835](https://doi.org/10.1109/EMBC.2016.7590835).
- Kong, W., Wang, L., Xu, S., Babiloni, F., Chen, H., 2019a. EEG fingerprints: phase synchronization of EEG signals as biomarker for subject identification. *IEEE Access* 7, 121165–121173. doi:[10.1109/ACCESS.2019.2931624](https://doi.org/10.1109/ACCESS.2019.2931624).
- Korczowski, L., Cederhout, M., Andreev, A., Cattan, G., Rodrigues, P.L.C., Gautheret, V., Congedo, M., 2019. Brain Invaders calibration-less P300-based BCI with modulation of flash duration Dataset (bi2015a). 10.5281/zenodo.3266930
- Krusienski, D.J., Grosse-Wentrup, M., Galán, F., Coyle, D., Miller, K.J., Forney, E., Anderson, C.W., 2011. Critical issues in state-of-the-art brain-computer interface signal processing. *J. Neural Eng.* doi:[10.1088/1741-2560/8/2/025002](https://doi.org/10.1088/1741-2560/8/2/025002).
- Kumar, M.G., Narayanan, S., Sur, M., Murthy, H.A., 2021. Evidence of task-independent person-specific signatures in EEG using subspace techniques. *IEEE Trans. Inf. Forensics Secur.* 16, 2856–2871. doi:[10.1109/TIFS.2021.3067998](https://doi.org/10.1109/TIFS.2021.3067998).
- La Rocca, D., Campisi, P., Scarano, G., 2013. On the repeatability of EEG features in a biometric recognition framework using a resting state protocol. In: Proceedings of the BIOSIGNALS - International Conference on Bio-inspired Systems and Signal Processing, pp. 419–428. doi:[10.5220/0004339104190428](https://doi.org/10.5220/0004339104190428).
- Langer, N., Ho, E.J., Alexander, L.M., Xu, H.Y., Jozanovic, R.K., Henin, S., Petroni, A., Cohen, S., Marcelle, E.T., Parra, L.C., Millham, M.P., Kelly, S.P., 2017. A resource for assessing information processing in the developing brain using EEG and eye tracking. *Sci. Data* 4, 1–20. doi:[10.1038/sdata.2017.40](https://doi.org/10.1038/sdata.2017.40).
- Lee, M.H., Kwon, O.Y., Kim, Y.J., Kim, H.K., Lee, Y.E., Williamson, J., Fazli, S., Lee, S.W., 2019. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *Gigascience* 8, 1–16. doi:[10.1093/gigascience/giz002](https://doi.org/10.1093/gigascience/giz002).
- Lotte, F., Guan, C., 2011. Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* 58, doi:[10.1109/TBME.2010.2082539](https://doi.org/10.1109/TBME.2010.2082539).
- Maiorana, E., 2021a. Learning deep features for task-independent EEG-based biometric verification. *Pattern Recognit. Lett.* 143, 122–129. doi:[10.1016/j.patrec.2021.01.004](https://doi.org/10.1016/j.patrec.2021.01.004).
- Maiorana, E., 2021b. Transfer learning for EEG-based biometric verification. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). BIBM 2021, pp. 3656–3661. doi:[10.1109/BIBM52615.2021.9669495](https://doi.org/10.1109/BIBM52615.2021.9669495).
- Maiorana, E., 2020. Deep learning for EEG-based biometric recognition. *Neurocomputing* 410, 374–386. doi:[10.1016/j.neucom.2020.06.009](https://doi.org/10.1016/j.neucom.2020.06.009).
- Maiorana, E., Campisi, P., 2018. Longitudinal evaluation of EEG-based biometric recognition. *IEEE Trans. Inf. Forensics Secur.* 13, 1123–1138. doi:[10.1109/TIFS.2017.2778010](https://doi.org/10.1109/TIFS.2017.2778010).
- Maiorana, E., La Rocca, D., Campisi, P., 2016. On the permanence of EEG signals for biometric recognition. *IEEE Trans. Inf. Forensics Secur.* 11, 163–175. doi:[10.1109/TIFS.2015.2481870](https://doi.org/10.1109/TIFS.2015.2481870).
- Marcel, S., Millan, J.R., 2007. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 743–748. doi:[10.1109/TPAMI.2007.1012](https://doi.org/10.1109/TPAMI.2007.1012).
- Olbrich, S., Conradi, J., 2016. Future of clinical EEG in psychiatric disorders: shifting the focus from diagnosis to the choice of optimal treatment. *Clin. Neurophysiol.* 127, 17–18. doi:[10.1016/J.CLINPH.2015.06.018](https://doi.org/10.1016/J.CLINPH.2015.06.018).
- Ozdenizci, O., Wang, Y., Koike-Akino, T., Erdoganmus, D., 2019. Adversarial deep learning in EEG biometrics. *IEEE Signal Process. Lett.* 26, 710–714. doi:[10.1109/LSP.2019.2906826](https://doi.org/10.1109/LSP.2019.2906826).
- Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R., 2019. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Sci. Data* 6, 1–5. doi:[10.1038/s41597-019-0104-8](https://doi.org/10.1038/s41597-019-0104-8), 2019 61.
- Pfurtscheller, G., Neuper, C., 2001. Motor imagery direct communication, 89, pp. 1123–1134. doi:[10.1109/5.939829](https://doi.org/10.1109/5.939829).
- Rodrigues, P.L.C., Jutten, C., Congedo, M., 2019. Riemannian procrustes analysis: transfer learning for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* 66, doi:[10.1109/TBME.2018.2889705](https://doi.org/10.1109/TBME.2018.2889705).
- Satti, A., Guan, C., Coyle, D., Prasad, G., 2010. A covariate shift minimization method to alleviate non-stationarity effects for an adaptive brain-computer interface. In: Proceedings of the International Conference on Pattern Recognition, pp. 105–108. doi:[10.1109/ICPR.2010.34](https://doi.org/10.1109/ICPR.2010.34).
- Seghier, M.L., Price, C.J., 2018. Interpreting and utilising intersubject variability in brain function. *Trends Cogn. Sci.* 22, 517–530. doi:[10.1016/j.tics.2018.03.003](https://doi.org/10.1016/j.tics.2018.03.003).
- Sui, J., Jiang, R., Bustillo, J., Calhoun, V., 2020. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. *Biol. Psychiatry* 1–31. doi:[10.1016/j.biopsych.2020.02.016](https://doi.org/10.1016/j.biopsych.2020.02.016).
- Sun, S., 2008. Multitask learning for EEG-based biometrics. In: Proceedings of the - International Conference on Pattern Recognition, pp. 1–4. doi:[10.1109/icpr.2008.4761865](https://doi.org/10.1109/icpr.2008.4761865).
- Snyder, A.Z., 1992. Steady-state vibration evoked potentials: description of technique and characterization of responses. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials* 84, 257–268. doi:[10.1016/0168-5597\(92\)90007-X](https://doi.org/10.1016/0168-5597(92)90007-X).
- Valizadeh, S.A., Riener, R., Elmer, S., Jänecke, L., 2019. Decrypting the electrophysiological individuality of the human brain: Identification of individuals based on resting-state EEG activity. *Neuroimage* 197, 470–481. doi:[10.1016/j.neuroimage.2019.04.005](https://doi.org/10.1016/j.neuroimage.2019.04.005).
- Wang, M., El-Fiqi, H., Hu, J., Abbas, H.A., 2019. Convolutional neural networks using dynamic functional connectivity for EEG-based person identification in diverse human states. *IEEE Trans. Inf. Forensics Secur.* 14, 3359–3372. doi:[10.1109/TIFS.2019.2916403](https://doi.org/10.1109/TIFS.2019.2916403).
- Wang, M., Hu, J., Abbas, H.A., 2020. BrainPrint: EEG biometric identification based on analyzing brain connectivity graphs. *Pattern Recognit* 105, doi:[10.1016/j.patcog.2020.107381](https://doi.org/10.1016/j.patcog.2020.107381).
- Wei, C.S., Keller, C.J., Li, J., Lin, Y.P., Nakanishi, M., Wagner, J., Wu, W., Zhang, Y., Jung, T.P., 2021. Editorial: inter- and Intra-subject variability in brain imaging and decoding. *Front. Comput. Neurosci.* 15, 102. doi:[10.3389/FNCOM.2021.791129](https://doi.org/10.3389/FNCOM.2021.791129).
- Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatranio, L.A., Robinson, C.J., Vaughan, T.M., 2000. Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* doi:[10.1109/TRE.2000.847807](https://doi.org/10.1109/TRE.2000.847807).
- Wolpaw, J.R., McFarland, D.J., Neat, G.W., Forneris, C.A., 1991. An EEG-based brain-computer interface for cursor control. *Electroencephalogr. Clin. Neurophysiol.* 78, 252–259. doi:[10.1016/0013-4694\(91\)90040-B](https://doi.org/10.1016/0013-4694(91)90040-B).
- Wu, Q., Yan, B., Zeng, Y., Zhang, C., Tong, L., 2018. Anti-deception: reliable EEG-based biometrics with real-time capability from the neural response of face rapid serial visual presentation. *Biomed. Eng. Online* 17, doi:[10.1186/S12938-018-0483-7](https://doi.org/10.1186/S12938-018-0483-7).
- Yang, S., Hoque, S., Deravi, F., 2022. Adaptive template enhancement for improved person recognition using small datasets. 10.48550/arXiv.2201.01218.
- Yingnan, S., Lo, F.P.W., Lo, B., 2019. EEG-based authentication with 1D-convolutional long short-term memory neural networks. *Expert Syst. with. Appl.* 125, 259–267. doi:[10.1016/j.eswa.2019.01.080](https://doi.org/10.1016/j.eswa.2019.01.080).
- Zeynali, M., Seyedarabi, H., 2019. EEG-based single-channel authentication systems with optimum electrode placement for different mental activities. *Biomed. J.* 42, 261–267. doi:[10.1016/j.bj.2019.03.005](https://doi.org/10.1016/j.bj.2019.03.005).
- Zheng, W.L., Liu, W., Lu, Y., Lu, B.L., Cichocki, A., 2019. EmotionMeter: a multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* 49, 1110–1122. doi:[10.1109/TCYB.2018.2797176](https://doi.org/10.1109/TCYB.2018.2797176).